# Single-molecule Fluorescence Spectroscopy of Protein Folding: Probing Denatured State Interactions and Folding Transitions

Dissertation
zur Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. Nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Sonja Müller-Späth
aus
Deutschland

Promotionskomitee

Prof. Dr. Benjamin Schuler (Vorsitz)
Prof. Dr. Andreas Plückthun
PD Dr. Ilian Jelezarov

Zürich 2010

# Summary

Protein folding is crucial for proteins to obtain their functional three-dimensional structure. It is known that different interactions, such as hydrophobic and electrostatic interactions as well as hydrogen bonds have an impact on folding, but their effects on the properties of the unfolded state are much less understood. Here, single molecule Förster resonance energy transfer (FRET) spectroscopy is used to monitor the dimensions of unfolded proteins and the differences between the folded and the unfolded state. This work consists of four topics: the first and second part treat the unfolded state collapse in protein folding and address the question which interactions are dominant; the third part presents a new approach to obtain dye distances from FRET measurements despite rotational restrictions of the dyes, and in the fourth part, the individual folding behavior of small immobilized proteins is investigated.

In the first part of the thesis, the collapse in intrinsically disordered proteins (IDPs) was observed dependent on temperature and on the charge content of the amino acid sequence. The comparison of the temperature-induced collapse of a two-state folder and an intrinsically disordered protein (IDP) revealed that the collapse is not only driven by hydrophobic interactions, but that additional forces contribute to the compaction in the denatured state. It is still not clear which interactions are the dominant ones. There are some indications that secondary structure is formed with increasing temperature that could lead to compaction by hydrogen bond formation.

In the second part of the thesis, one globular protein and two IDPs with different charge content were used to investigate systematically the impact of charges in the amino acid chain on the dimensions of the unfolded state and its denaturant-induced collapse. It was shown that the collapse is highly dependent on the average charge density and therefore dominated by electrostatic forces. A polyampholyte model was applied to the data and was shown to describe the attractive and repulsive contributions that account for the final protein dimensions. These findings are important steps towards the ultimate goal of quantifying the different contributions that influence unfolded state behavior. Additionally, information about the dimensions of IDPs and their strong salt dependence at low ionic strength was obtained. Consequences will be especially interesting with respect to the cellular function of IDPs.

In the third part of the thesis, a new approach to obtain distance information from FRET measurements with restricted dye rotation is presented. To determine dye distances in constrained environments, as in the case of a protein bound to the molecular chaperone GroEL, orientation factors of the dye have to be taken into account. By combining single molecule FRET and time-resolved fluorescence anisotropy decays, distance information can be obtained. In the case of

rhodanese, the results implicate that the protein binds to the chaperone as a partially structured folding intermediate.

In the fourth part of the thesis, the folding behavior of single immobilized proteins was investigated. For this approach, proteins were encapsulated in lipid vesicles and bound to a surface. These immobilization conditions were chosen because it is expected that the proteins can still diffuse and fold and unfold undisturbed as they can in free solution, but are confined in the observation plane of the single molecule instrument. However, since the recorded trajectories show very heterogeneous signal and are probably dominated by photophysics and photochemistry, it was not possible to determine the kinetics of single immobilized cold shock proteins. Only very few events could be assigned to folding or unfolding after a comprehensive analysis was performed to distinguish folding signal from signal that is due to dye properties, such as bleaching and darkstates. The methods developed in the course of this project will be important for future work on single immobilized protein molecules.

# Zusammenfassung

Die Faltung von Proteinen ist entscheidend für ihre funktionelle dreidimensionale Struktur. Es ist zwar bekannt, dass verschiedene Wechselwirkungen, wie hydrophobe und elektrostatische Wechselwirkungen, sowie Wasserstoffbrücken-Bindungen an der Proteinfaltung beteiligt sind, allerdings ist ihre Auswirkung auf den entfalteten Zustand noch nicht vollständig verstanden. In dieser Arbeit werden mit Hilfe von Einzelmolekülfluoreszenzspektroskopie, insbesondere unter Nutzung von Resonanzenergietransfer nach Förster (FRET) die räumliche Ausdehnung des entfalteten Zustands sowie die Unterschiede zwischen ge- und entfaltetem Zustand untersucht. Die Arbeit besteht aus vier Abschnitten: Im ersten und zweiten Teil wird der Proteinkollaps im entfalteten Zustand behandelt und die Fragestellung bearbeitet, welche Wechselwirkungen in Proteinen dominieren. Im dritten Teil wird ein neuer Ansatz gezeigt, der es ermöglicht, Farbstoffabstände aus FRET-Messungen zu erhalten, auch wenn die freie Rotation der Fluoreszenzfarbstoffe eingeschränkt ist. Im vierten Teil der Arbeit wird das Faltungsverhalten einzelner immobilisierter Proteine beschrieben.

Im ersten Teil der Arbeit wurde der Proteinkollaps von intrinsisch ungeordneten Proteinen (englisch *intrinsically disordered proteins*, IDPs) unter dem Einfluss von Temperatur und dem Ladungsgehalt der Aminosäuresequenz beobachtet. Der Vergleich des temperaturinduzierten Kollapses im entfalteten Zustand eines Zwei-Zustands-Falters und eines IDPs ergab, dass der Kollaps nicht nur durch hydrophobe Wechselwirkungen verursacht wird, sondern dass darüber hinaus weitere Kräfte zur Kompaktierung beitragen. Es ist noch nicht bekannt, welche Wechselwirkungen dabei dominieren. Es gibt einige Hinweise, dass mit steigender Temperatur Sekundärstruktur entsteht, die zu einer Kompaktierung durch die Bildung von Wasserstoffbrücken-Bindungen führen könnte.

Im zweiten Teil der Arbeit wurden ein globuläres Protein und zwei IDPs mit unterschiedlichem Ladungsgehalt verwendet, um den Einfluss von Ladungen in der Aminosäure-Kette auf die räumliche Ausdehnung des entfalteten Zustands und den Denaturans-induzierten Kollaps systematisch zu untersuchen. Es konnte gezeigt werden, dass der Kollaps stark von der durchschnittlichen Ladungsdichte abhängig ist und deshalb von elektrostatischen Wechselwirkungen dominiert wird. Ausserdem wurde gezeigt, dass ein Polyampholyt-Modell, das auf die Daten angewendet wurde, die anziehenden und abstossenden Kräfte beschreiben kann, die für die endgültigen Proteindimensionen verantwortlich sind. Diese Erkenntnisse sind ein wichtiger Schritt zum Ziel, die verschiedenen Wechselwirkungen, die das Verhalten des entfalteten Zustands beeinflussen, zu quantifizieren. Darüber hinaus wurden Kenntnisse über die Dimensionen von IDPs und deren starke Salzabhängigkeit bei niedrigen Ionenstärken gewonnen.

Die Tragweite dieser Erkenntnisse ist insbesondere im Hinblick auf die Funktion der IDPs interessant.

Im dritten Teil der Arbeit wird eine neue Methode vorgestellt, die es ermöglicht, trotz eingeschränkter Farbstoffrotation Abstandsinformationen zu gewinnen. Um Farbstoffabstände in räumlich eingeschränkten Umgebungen messen zu können, wie im Fall von Protein, das an das Molekulare Chaperon GroEL gebunden ist, muss die eingeschränkte Rotationsfreiheit der Farbstoffe berücksichtigt werden. Durch die Verwendung von Einzelmolekül-FRET und zeitaufgelösten Fluoreszenz-Anisotropie-Zerfällen konnten Abstandsinformationen ermittelt werden. Bei dem Protein Rhodanese implizieren die Ergebnisse, dass das Protein als teilweise strukturiertes Faltungsintermediat an das Chaperon bindet.

Im vierten Teil der Arbeit wurde das Faltungsverhalten von einzelnen immobilisierten Proteinen untersucht. Hierfür wurden Proteine in Lipidvesikel eingeschlossen und auf eine Oberfläche gebunden. Es wurden Immobilisierungsbedingungen gewählt, bei denen man davon ausgeht, dass die Proteine zwar unbeeinträchtigt wie in freier Lösung diffundieren, falten und entfalten können, die Aufenthaltsmöglichkeiten aber beschränkt sind auf die Beobachtungsebene des Einzelmolekül-Instruments. Allerdings war es nicht möglich, die kinetischen Eigenschaften von einzelnen immobilisierten Kälteschockproteinen zu bestimmen, weil die gemessenen Fluoreszenztrajektorien ein sehr heterogenes Signal zeigten, das wahrscheinlich von Photophysik und Photochemie dominiert wird. Nur einige wenige Ereignisse konnten der Faltung oder Entfaltung zugeordnet werden, nachdem eine umfassende Analyse durchgeführt wurde, mit der Faltungssignale von anderen Signalen unterschieden werden konnten, die abhängig von Farbstoffeigenschaften wie Bleichen und der Bildung von Dunkelzuständen sind. Die im Zusammenhang mit dem Projekt entwickelten Methoden werden für weiterführende Projekte mit einzelnen immobilisierten Proteinen von Bedeutung sein.

# Publications

**Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy.**

F. Hillger, D. Hänni, D. Nettels, S. Geister, M. Grandin, M. Textor, & B. Schuler

*Angew Chem Int Ed Engl,* **2008***, 47*, 6184-6188

**Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins.**

D. Nettels, , S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, & , B. Schuler

*Proc Natl Acad Sci U S A,* **2009**, *106*(49):20740-20745

**Charge interactions can dominate the dimensions of intrinsically disordered proteins.**

S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, B. Schuler

*Proc Natl Acad Sci U S A,* **2010**, early edition 1-6

# Contents

# Introduction

Proteins are involved in very different processes in living organisms such as in metabolism or regulatory processes in transcription and translation. Structural proteins constitute the cytoskeleton or support the stability of organisms in the extracellular space. To be functional, most proteins have to adopt a specifically folded structure. For these diverse functions, proteins of various structures are needed. The native state of the protein is a unique conformation that is stable only under a limited range of conditions, which are typically close to physiological. If the conditions are changed to non-physiological, by the addition of a chemical denaturant, such as guanidinium chloride (GdmCl) or urea, by heating, or by moving to extreme pH values in the solution, the protein is denatured. In contrast to the native state, the denatured state is an ensemble of many conformations.

For many proteins, the unfolding process is reversible, and they fold back spontaneously to their native structure if the surrounding solution is adjusted to physiological conditions (1). Since this refolding is possible, the information to adopt the native structure must be included in the amino acid sequence. The different amino acids have different hydrophobicity, polarity, and charge, and are therefore responsible for different interactions within the protein chain. These interactions are, according to the features of the amino acids, hydrophobic and electrostatic interactions, as well as hydrogen bonds that are formed between polar residues, and they drive the folding process towards the native structure. However, the general problem of protein folding, and the question which interactions are dominating the folding process, are still not solved.

## *From the unfolded to the native state*

Different proteins fold according to different mechanisms, but in general, the folding and unfolding pathways between the different states can be considered as diffusive paths on an energy landscape. The native state is located at a coordinate where the free energy is lowest (2). Because the unfolded state is an ensemble of unfolded conformations, it can be located at different coordinates on the energy landscape. Therefore, multiple pathways to the native state exist for each protein. Depending on the mechanism, proteins can either fold directly or via intermediates, which are local energy minima in the folding landscape. Several multi-domain proteins are even dependent on the aid of chaperones in a cellular environment, as will be described below.

In the simplest case, proteins fold according to the two-state model, as many small single domain proteins do. In this model, only two thermodynamic states are occupied: the native and the denatured state, but no intermediates. The states are separated by a free energy barrier, which has to be overcome each time a protein folds or unfolds. The proteins remain in either the unfolded or the folded state for extended periods of time, with intermittent rapid jumps across the barrier.

Interestingly, many proteins have been shown to exhibit a compaction of the unfolded state at equilibrium conditions when the denaturant concentration is reduced (3-8) This collapse is expected to reflect the sum of different interactions that drive a protein to a compact state, which can support the transition to the folded state (9). However, it still has to be clarified which interactions are involved, and to which extent they contribute to the collapse. The second focus of the thesis was to find out whether the collapse is a general behavior for all proteins. Therefore, intrinsically disordered proteins (IDPs) were chosen for an investigation of their collapse behavior. With their high fraction of charged amino acids and low content of hydrophobic amino acids, the relative contributions of hydrophobic and electrostatic interactions should differ. For these projects, prothymosin α (ProTα) and the N-terminal domain of HIV-1 integrase (IN), two proteins with very different properties concerning mean charge and hydrophobicity, were chosen. The aim was to quantify the degree of collapse and investigate how much it depends on the mean charge and hydrophobicity of the protein (Chapter 1 and 2).

In contrast to small single domain proteins, many other proteins are not able to fold directly to the native conformation. These consist typically of more than one domain and then fold via intermediates. These proteins can show complex kinetics dependent on how many domains they have and whether they differ in kinetic stability. Many larger multi domain proteins fold via several intermediates, which slow down the folding and are therefore prone to aggregation. Thus, they are often dependent on the assistance of molecular chaperones, especially at the high protein concentrations in the cell. A well-studied example is the bacterial GroEL/GroES system of the Hsp60 class. This heptameric complex assists its substrate proteins during folding by binding non-native proteins in its cavity, allowing the proteins to fold, and releasing them. Complete folding can require several cycles of binding and release. Until now, not much has been known about substrate protein conformations in the chaperone during folding. Bovine rhodanese is a substrate of GroEL/GroES, and its interaction with the chaperone was investigated with single molecule FRET (Chapter 3).

One further objective of this thesis was to observe the frequencies of folding and unfolding transitions of single immobilized molecules (Chapter 4). For the cold shock proteins (Csps) of *Thermotoga maritima* and *Bacillus caldolyticus* used in this work, it has been shown in ensemble (10-11) and in single molecule experiments (8) that they fold and unfold in a two-state manner. From ensemble measurements, it is known that the proteins cross this barrier with equal folding and unfolding rate at the transition midpoint. By monitoring the frequency of transitions, the goal was to investigate whether the stochastic folding of a single molecule follows the folding dynamics known from ensemble measurements, or whether there is a broader distribution of folding rates. A larger distribution would be the result of many different individual dynamic behaviors.

## *Different interactions contribute to collapse*

There are different interactions that are assumed to drive folding to the native structure: Formation of hydrogen bonds, hydrophobic interactions, and electrostatic interactions are assumed to contribute. Although there have been many attempts during the last 60 years to approach this problem, it is still not solved.

Initially, Pauling and coworkers suggested in 1951 that the driving force in protein folding would be hydrogen bonding (12). Schellman confirmed this hypothesis in 1955 by finding that the intra-peptide hydrogen bonds are favored over hydrogen bonds with water (13). Later, hydrogen bonds were no longer believed to contribute strongly to folding, until from 1991 on, opposing results and interpretations revived the discussion. Hydrogen bonds were suggested again to stabilize proteins (14-17), whereas in other works it was argued that hydrogen bonds are destabilizing (18-19) or partially stabilizing and destabilizing (20). This was also found in recent work by Kiefhaber and coworkers (21): in experiments on model peptides, a decrease in chain dimension in water, compared to that in GdmCl, was attributed to the formation of hydrogen bonds. End-to-end diffusion constants under native conditions were slowed down, but loop formation was accelerated due to shorter diffusion paths. Therefore, it was concluded that hydrogen bonds on the one hand favor protein structures that are more compact and on the other hand decrease the diffusion time due to non-specific bonds.

The hydrophobic effect was first suggested to be responsible for folding by Kauzmann in 1959 (22). His opinion was supported by studies of different model systems (23-24) and has been accepted for many years. Additionally, in more recent work of Muñoz and coworkers, the collapse was assigned to hydrophobic interactions (25).

Another potential factor with influence on the compaction of the unfolded state are electrostatic interactions that depend on the number of charges in the protein. Uversky *et al.* (26) found that IDPs can be distinguished from globular proteins by their differences in charge and hydrophobicity in the amino acid sequence, suggesting that in globular proteins, attractive forces (the hydrophobic effect and opposite charges) are stronger than repulsive forces (like electrostatic charges), whereas in IDPs hydrophobic interactions are not strong enough to compensate repulsion from charges. A first model to calculate electrostatic forces in the unfolded protein, assuming a Gaussian chain, was developed by Zhou (27). Also in another work, electrostatic interactions were shown to counteract the compaction of the unfolded state in a small protein (28).

Despite many approaches assessing the collapse, it is still a subject of speculation to what extent these three major interactions are contributing. For a more detailed summary of the historic events, see works by Bolen, Dill, and Rose (17, 19, 29).

## *Intrinsically disordered proteins*

IDPs, also called “natively unstructured” or “natively unfolded” proteins, are unfolded under near-physiological conditions *in vitro*. The absence of structure is assumed to be due to their high net charge and polarity that prevents the formation of a hydrophobic core typical of globular proteins (26). The difference in amino acid composition was shown by plotting the mean net charge against the mean hydrophobicity (26), which was calculated according to values from Kyte and Doolittle (30). There are two areas in the plot separated by a border dividing globular proteins from IDPs. Many of the IDPs can fold spontaneously upon binding to their ligands. One potential advantage of the intrinsic lack of structure and this function-related disorder-to-order transitions is the coupling of high specificity with low affinity (31-34). In this case, the energy needed to organize the disordered region is compensated by the free energy arising from forming contacts with binding partners. Due to the flexibility of the protein, specific binding to multiple ligands is enabled (32, 35-36).

The flexible, unstructured proteins or domains can also have linker function and connect globular or transmembrane domains in many eukaryotic proteins. Such long linker regions facilitate for instance the assembly of transcriptional complexes, for which the interaction with several regulatory factors is necessary (37). These properties may be highly useful in transcription and signaling regulation, as well as in cell cycle control (38). Generally, unstructured proteins are prone to degradation, but the coupled binding and folding can protect proteins from proteolysis. It is assumed that regions with exposed hydrophobic amino acids are targeted by degradation mechanisms of the cell (39), but linker sequences require long-term stability that can be ensured by their low content of hydrophobic residues (37).

## *Exploring the denatured state with single molecule FRET spectroscopy*

Ensemble experiments have been used extensively for many years to observe protein folding, and many proteins were analyzed thermodynamically. For most methods, the signal obtained is an average of the folded and the unfolded state or even of intermediates. Only under few conditions (e.g. native or completely denaturing), one population is dominating the average signal, and information about a specific state can be obtained. By overcoming these limitations, single molecule experiments can be used to probe the denatured state even in the presence of the native state. These experiments reveal the distribution of the denatured conformations and will help to gain insight into protein folding.

One method frequently used in single molecule experiments is Förster resonance energy transfer (FRET). This method allows the direct observation of the folding of single proteins and the differentiation between populations that is not possible with ensemble methods. For single molecule FRET, two fluorescent dyes - acceptor and donor - are covalently attached to the protein, the donor is excited, and the fluorescence light is collected, split into donor and acceptor photons, and detected separately. After the signal is corrected for quantum yields of the dyes, spectral cross-talk, and direct excitation of the acceptor, the energy transfer efficiency can be calculated for every molecule. The transfer efficiency $E$ from the donor to the acceptor is proportional to the inverse sixth power of the distance between the dyes. From the mean energy transfer efficiency, $\langle E \rangle$, of the transfer efficiency peak of the unfolded state, the corresponding distance distribution between the dyes can be calculated, given the Förster radius of the dye pair and a suitable distance distribution function. The shape of the distance probability distribution can be accessed directly only by fluorescence lifetime measurements. Fluorescence lifetime decays of unfolded Csp$Tm$, for example, are fit well with the distance probability distributions of a Gaussian chain (40). The agreement between distance results obtained from lifetime decays and fluorescence intensity experiments additionally supports a Gaussian chain-like behavior in this case.

By applying this method, the equilibrium collapse of the denatured state of Csp$Tm$ was investigated, which had been inaccessible with ensemble measurements. The collapse could be observed in single molecule experiments even under near native conditions (3, 40). Whereas the denatured state of the proteins investigated does not exist under fully native conditions, it exists for destabilized proteins and for IDPs and could be characterized (41-42) (Chapter 2). Additionally, partial structure formation upon binding of zinc by the IDPs could be excluded by using single molecule experiments. Furthermore, the experiments allowed us to investigate whether there are additional populations or just one population with different energy transfer efficiency at different denaturant or salt concentrations.

## Observing the folding and unfolding of immobilized molecules via FRET

In two-state folding, protein molecules remain in either the unfolded or the folded state for extended periods of time, with intermittent rapid jumps across the free energy barrier. Measuring fluorescence trajectories of single proteins during folding potentially provides the opportunity to observe folding dynamics in the time range of seconds and may complement methods such as molecular dynamic simulations that give access to shorter time ranges in the order of nano- and microseconds.

The prerequisite for observing the folding behavior of single molecules over a longer time is to hold them in place during the observation time. Different techniques to immobilize molecules on the surfaces have been used. Proteins (43-45) and rybozymes (46) have been encapsulated in lipid vesicles tethered to a supported lipid bilayer on glass. In a different immobilization technique used by Kuzmenkina *et al.* (47), proteins were directly tethered to a PEG-biotinylated surface. For this work, star-shaped, strongly cross-linked PEG was used, which was developed specifically to minimize non-specific adsorption. In a recent work, Chung *et al.* (48) used brush-like PEG-coated glass cover slides and biotin-avidin linkage to immobilize proteins.

Immobilized FRET-labeled molecules can either be observed one by one by confocal microscopy, or by total internal reflection fluorescence (TIRF), where the signal of many molecules can be recorded at a time. Both methods result in fluorescence trajectories that should reflect dye to dye distances within the protein. In two-state folding, only two observable transfer efficiencies are expected: one for the folded and one for the unfolded state. But there are also reports of folding trajectories of individual proteins observed with FRET or atomic force microscopy (AFM) (43, 49) with an unexpected degree of complexity rather than simple bi-stable behavior. This is in contrast to the simplicity of the kinetic model of two-state folding and raises the question whether the simple kinetic properties derived from classical experiments on large ensembles of molecules are always reflected in the folding paths taken by individual proteins (43, 49-50).

Immobilization experiments are still accompanied by some difficulties that have to be solved. Some of the observed complexity could be caused by factors such as unspecific binding of the sample or the dyes to the surface, which is strongly dependent on the chosen immobilization system, by the choice of dyes, which are at risk of undergoing transitions to dark states, interaction between biomolecule and dyes, and subjectivity involved in trajectory selection, or the criteria for the automated trajectory analysis. A general problem in immobilization experiments is that the sample is illuminated for long times, which leads to bleaching of the fluorescent dyes. For diffusion experiments, most of these factors are not limiting because the samples are exposed to the laser for shorter times and the molecules are not influenced by surface interactions. If the fluorophores are bleached, they either are not visible because the acceptor cannot be excited if the donor is damaged, or the event is visible as a separate donor-only population if an active acceptor is missing.

Another limitation in single molecule FRET experiments are the Förster radii of the dye pairs suitable for single molecule detection in the range of 5 nm, which are thus most suitable to measure distances in the range of 3 to 8 nm. To extend the distance range, approaches using multiple dyes have been developed and applied (51-53). Even unidirectional five-dye FRET was done with immobilized “photonic wires” made from DNA, and FRET could be detected of over a distance of more than 13 nm (54). With multiple-dye FRET, it is possible to probe several distances in a molecule at a time (or within multiple component complexes). Applied to proteins, this would mean that distance information gained from three different protein variants can then be obtained by one variant with three labels. This might be of special interest for subdomain folding and has the advantage that different intra-chain distances can be observed simultaneously and can be correlated with each other.

There are many approaches to address the challenges described in the case of immobilized molecules. Because one major challenge in the immobilization experiments is the limited photo-stability of the dyes during long illumination, several methods have been developed to overcome this issue. For example, time-lapse excitation has been used (47), which is, however, only suitable for proteins with low folding and unfolding rates because the time resolution is decreased and folding information of proteins with fast rates would be lost. Other possibilities are the use of protecting additives that prevent dyes from bleaching (55-57). Bleaching of the dyes is especially disadvantageous if the protein folds with a low folding rate and only few or no transitions can be observed until the dyes bleach. The work presented here indicates that some of the complications in single molecule folding experiments on immobilized proteins are still very difficult to overcome and can interfere strongly with the identification of conformational transitions.

# 1. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins

*Proc Natl Acad Sci U S A*, **2009**, 106(49):20740-20745

# Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins

**Daniel Nettels[a], Sonja Müller-Späth[a], Frank Küster[a], Hagen Hofmann[a], Dominik Haenni[a], Stefan Rüegger[a], Luc Reymond[a], Armin Hoffmann[a], Jan Kubelka[b], Benjamin Heinz[c], Klaus Gast[c], Robert B. Best[d], and Benjamin Schuler[a,1]**

[a]Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; [b]Department of Chemistry, University of Wyoming, Laramie, WY 82071; [c]Physikalische Biochemie, Universität Potsdam, 14476 Potsdam, Germany; and [d]Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, United Kingdom

Edited by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved September 25, 2009 (received for review March 6, 2009)

**We used single-molecule FRET in combination with other biophysical methods and molecular simulations to investigate the effect of temperature on the dimensions of unfolded proteins. With single-molecule FRET, this question can be addressed even under near-native conditions, where most molecules are folded, allowing us to probe a wide range of denaturant concentrations and temperatures. We find a compaction of the unfolded state of a small cold shock protein with increasing temperature in both the presence and the absence of denaturant, with good agreement between the results from single-molecule FRET and dynamic light scattering. Although dissociation of denaturant from the polypeptide chain with increasing temperature accounts for part of the compaction, the results indicate an important role for additional temperature-dependent interactions within the unfolded chain. The observation of a collapse of a similar extent in the extremely hydrophilic, intrinsically disordered protein prothymosin $\alpha$ suggests that the hydrophobic effect is not the sole source of the underlying interactions. Circular dichroism spectroscopy and replica exchange molecular dynamics simulations in explicit water show changes in secondary structure content with increasing temperature and suggest a contribution of intramolecular hydrogen bonding to unfolded state collapse.**

FRET | polymer | protein folding | secondary structure | chain dimensions

There is an increasing interest in the properties of unfolded proteins and their roles in the folding and cellular functions of proteins. A key motivation is that many proteins are marginally stable and only fold in the presence of their ligands or binding partners, opening new regulatory possibilities (1, 2). An important reason for recent progress is the growing availability of methods that provide structural information on these conformationally heterogeneous systems, such as NMR (3), scattering methods (4, 5), and single-molecule FRET (6–8). Although NMR provides mostly local details, small-angle X-ray scattering (SAXS), dynamic light scattering (DLS), and single-molecule FRET provide overall hydrodynamic or long-range distance information. An important advantage of single-molecule FRET is the separation of folded and unfolded subpopulations (9). As a result, unfolded state properties can be investigated even in the presence of folded molecules (i.e., under near-native conditions, which are physiologically most relevant). This advance has led to the observation of a continuous collapse of the unfolded state with decreasing denaturant concentrations (10), a behavior that now has been demonstrated for a large number of proteins (11–18) and peptides (19). Recent advances in the application of theoretical models have led to a quantitative description of this unfolded state collapse in terms of polymer-physical concepts (15, 20–23). Such chain compaction also has been demonstrated to result in increased internal friction and a slowdown of intramolecular dynamics of the polypeptide (19, 26), which can affect the kinetics of protein folding (27). However, it is still unclear which interactions drive unfolded state collapse.

An experimental variable that can provide more information about the nature of these interactions is temperature. The temperature dependence of protein collapse also has a substantial bearing on several fundamental aspects in protein folding, especially on the degree of energetic frustration and, correspondingly, on folding kinetics (28–33). For simple polymers and polymer models that do not involve a temperature-dependent interaction energy, conformational entropy favors open conformations, leading to chain expansion with increasing temperature (34), an effect that is assumed frequently to be dominant also for unfolded proteins. Although none of the sparse experimental results on this topic show such an expansion, they exhibit substantial variation, ranging from the absence of any detectable temperature dependence (35, 36) to slight (37, 38) and stronger (39) collapse, with some inconsistencies between measurements even on the same protein (35, 37, 40), calling for a systematic investigation of this issue. Ensemble investigations are limited largely to highly unfolding conditions to exclude interference from the signal of folded molecules and to minimize aggregation. Although these limitations can be avoided in single-molecule FRET experiments (6–8), the analysis of changes in the conformational distribution of unfolded proteins has remained largely limited to the influence of denaturants. Possible reasons may be insufficient precision of previous single-molecule experiments or difficulties in accurate temperature control of the confocal optics used over a sufficiently broad temperature range. Here, we overcome these complications (Fig. S1) and use the combination of single-molecule FRET with several other biophysical methods to investigate the effect of temperature on unfolded state structure and dimensions over a wide range of denaturant concentrations.

## Results

**Single-Molecule FRET Experiments.** We first used a variant of the small cold shock protein Csp*Tm* (10, 16, 26, 41) labeled with donor (Alexa Fluor 488) and acceptor (Alexa Fluor 594) dyes close to its termini (10, 16). The efficiency of energy transfer between the dyes upon donor excitation was determined from photon bursts originating from individual molecules freely diffusing through the focal spot of the laser beam as $E = n_A/(n_A + n_D)$, where $n_A$ and $n_D$ are, respectively, the number of acceptor and donor photons emitted by the molecule (including corrections; see *SI Materials and Methods*). A histogram from a large number of such events shows distinct maxima corresponding to the subpopulations present in the sample (Fig. 1). The peak at high $E$ corresponds to folded molecules, and the peak at

---

Author contributions: D.N., S.M.-S., R.B.B., and B.S. designed research; D.N., S.M.-S., F.K., H.H., D.H., S.R., L.R., A.H., J.K., B.H., K.G., R.B.B., and B.S. performed research; D.N., S.M.-S., F.K., D.H., S.R., L.R., A.H., J.K., B.H., K.G., R.B.B., and B.S. contributed new reagents/analytic tools; D.N., S.M.-S., F.K., H.H., B.H., K.G., R.B.B., and B.S. analyzed data; and D.N., S.M.-S., K.G., R.B.B., and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: schuler@bioc.uzh.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900622106/DCSupplemental.

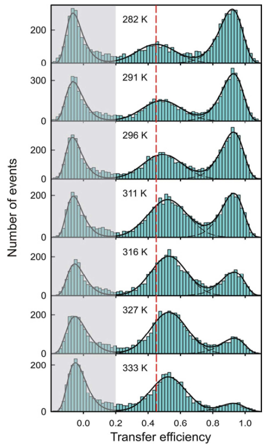

**Fig. 1.** Single-molecule FRET efficiency histograms of Csp*Tm* show thermal denaturation and a temperature-induced collapse of the unfolded protein (1.0 M guanidinium chloride). The peak at $E \approx 0.9$ corresponds to folded molecules; the peak at intermediate $E$ corresponds to unfolded molecules. The peak at $E \approx 0$ (shaded) originates from molecules with an inactive acceptor. To determine mean transfer efficiencies, the unfolded peak was fit to a normal distribution, and the other two peaks were fit to log-normal distributions (black lines).

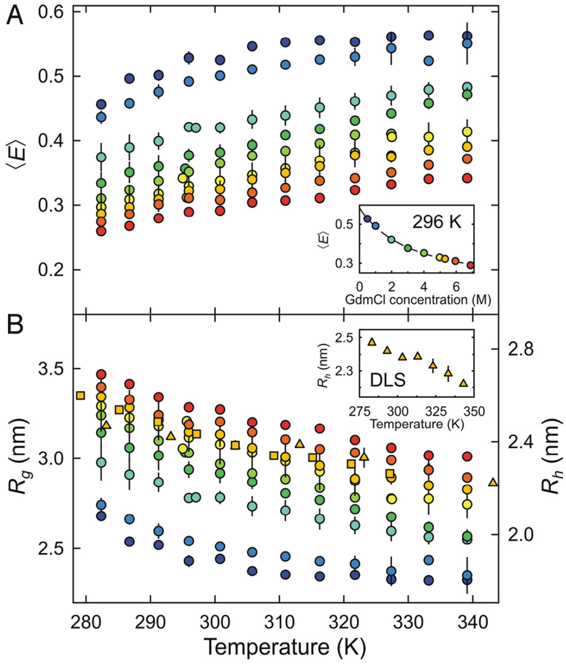

**Fig. 2.** Temperature dependence of (*A*) the mean transfer efficiency $\langle E\rangle$ and (*B*) the radius of gyration ($R_g$) of unfolded Csp*Tm* at different guanidinium chloride (GdmCl) concentrations (0.49, 1.0, 2.0, 3.0, 4.0, 5.0, 5.3, 6.0, and 6.9 M, from dark blue to red). Values from single-molecule measurements are shown as circles. The $R_g$ values from ensemble fluorescence lifetime measurements in 5.3 M GdmCl are shown as orange squares, and the hydrodynamic radii ($R_h$) from dynamic light scattering (DLS) measurements in 5.3 M GdmCl are shown as orange triangles (right axis). The mean transfer efficiencies at 296 K (*inset* in *A*) illustrate the denaturant-dependent unfolded state collapse. For clarity, the inset in (*B*) shows the DLS results. Error bars indicate standard deviations estimated from two or three independent measurements for the cases where several measurements are available.

intermediate $E$ corresponds to unfolded molecules. Fig. 1 shows an example of a temperature-dependent single-molecule FRET measurement at a guanidinium chloride (GdmCl) concentration of 1.0 M. With increasing temperature, we observe a decrease in the folded population and a concomitant increase in the unfolded population, as expected from the thermal denaturation of a two-state protein. However, we also observe an increase in the mean transfer efficiency of the unfolded subpopulation, $\langle E\rangle$, with temperature, corresponding to a collapse of the chain. Fig. 2*A* shows that this increase in $\langle E\rangle$ with temperature occurs at all GdmCl concentrations. The small change in transfer efficiency with temperature is probably the reason why we have not been able to detect a significant amplitude from unfolded state collapse in laser temperature-jump experiments, similar to experiments with a closely related cold shock protein (42). The transfer efficiency of unfolded Csp*Tm* at a constant temperature decreases with increasing GdmCl concentrations (Fig. 2*A Inset*), reflecting the well-established denaturant-induced expansion of the polypeptide (10–19).

For an analysis in terms of chain dimensions, the transfer efficiencies need to be converted to a measure of molecular size. Previous results indicate that on the length scales probed here, the distance distributions within unfolded Csp*Tm* can be approximated by those of a Gaussian chain (16). Taking into account the distance distribution $P(r)$, $\langle E\rangle$ in the unfolded state can be expressed as (10, 16)

$$\langle E\rangle = \int_a^{l_c} E(r)P(r)dr \Bigg/ \int_a^{l_c} P(r)dr, \text{ with } E(r) = 1/(1+(r/R_0)^6), \qquad \textbf{[1a,b]}$$

$$\text{and } P(r) = 4\pi r^2\left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2}\exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right), \qquad \textbf{[1c]}$$

where $r$ is the distance between donor and acceptor and $R_0$ is the Förster radius. Given a value of $\langle E\rangle$ from the unfolded subpopulation, the mean-squared end-to-end distance $\langle r^2\rangle$ of the chain thus can be calculated numerically. To provide an intuitively accessible quantity, we plot our data in terms of the radius of gyration (Fig. 2*B*), which for a Gaussian chain is given by $R_g^2 = \langle r^2\rangle/6$ (34). Choosing a different polymer model for the analysis (e.g., that of a worm-like chain) does not affect our conclusions because of the similar shapes of the distance distributions in the range of persistence lengths relevant here (22). Note also that the effect of chain dynamics on the observed transfer efficiencies (43) does not affect our analysis significantly (Fig. S2).

An ideal complementation of the distance information from FRET efficiency histograms, where the distance information is averaged over the duration of the fluorescence bursts, is the analysis of fluorescence intensity decays, which occur on a time scale of a few nanoseconds, much shorter than the reconfiguration time of the chain (26), and thus provide direct information about $P(r)$ (12, 16, 44). Additionally, fluorescence lifetimes are less susceptible to inaccuracies in instrument calibration. We thus performed ensemble time-correlated single photon counting experiments in 5.3 M GdmCl (Fig. 2*B*). The good agreement of the resulting values of $R_g$ with the single-molecule measurements (Fig. 2*B*) confirms the accuracy of the correction factors (*SI Materials and Methods*) used for the determination of transfer efficiencies and distances from the $E$ histograms. Additionally, we performed DLS experiments under identical conditions (5.3 M GdmCl) with unlabeled Csp*Tm*. Despite the less extensive statistics compared with the FRET experiments and some uncertainty from the slight protein concentration depen-

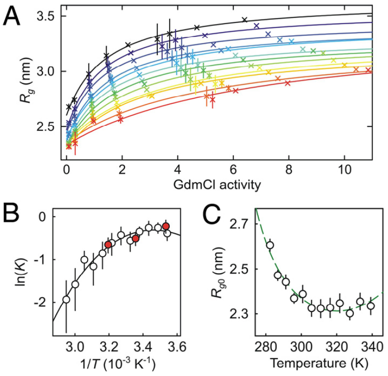

**Fig. 3.** Thermodynamic analysis. (*A*) The radius of gyration ($R_g$) of unfolded Csp*Tm* as a function of guanidinium chloride (GdmCl) activity at different temperatures (282, 287, 291, 296, 301, 306, 311, 316, 322, 327, 333, 339 K, from black to red crosses) with fits using Eq. **2** (lines). Error bars indicate standard deviations estimated from two to three independent measurements for the cases where several measurements are available. (*B*) The van't Hoff plot of the effective binding constants from the fits in *A* (open circles) compared with the values from calorimetric measurements taken from ref. 45 (red circles). (*C*) The radii of gyration at zero GdmCl activity ($R_{g0}$) from the fits in *A* show the temperature-induced collapse in the absence of denaturant. The dashed line is a fit to Eq. **3** ($T_0 = 295$ K, $\Delta H_0 = -1.3 \pm 0.1$ kJ mol$^{-1}$, $\Delta S_0 = -8.5 \pm 0.4$ J mol$^{-1}$ K$^{-1}$, $\Delta C_p = 50 \pm 9$ J mol$^{-1}$ K$^{-1}$). Error bars in *B* and *C* represent standard deviations taken from the covariance matrices of the fits in *A*.

dence of DLS measurements, the hydrodynamic radius $R_h$ clearly decreases with temperature by ≈10 ± 3% between 10 and 70 °C, compared with a decrease of $R_g$ by ≈15 ± 2% from FRET. The ensemble experiments under completely unfolding conditions thus agree with the single-molecule results, confirm the accurate calibration of the single-molecule instrument, and exclude a strong influence of the fluorophores on chain collapse. The agreement of single-molecule and DLS experiments suggests that the collapse of the unfolded state is a global process that affects the entire chain. Experiments with Csp*Tm* variants labeled at different positions (Fig. S3) support this conclusion.

**Thermodynamic Analysis.** In a first step toward the identification of the factors contributing to the temperature-induced chain collapse, we compare our results with the thorough calorimetric measurements of Makhatadze and Privalov on the interaction of denaturants with unfolded polypeptides (45). With increasing temperature, they found a decrease in the effective binding constants of urea and GdmCl to apo-cytochrome *c*, reduced and carboxymethylated ribonuclease, and lysozyme. The resulting temperature-induced dissociation of denaturant from the polypeptide is expected to lead to a collapse of the chain. Could this effect be sufficient to account for our observations?

A quantitative comparison requires the conversion of our results (Fig. 2) from GdmCl concentrations to thermodynamic activities *a* (45) to obtain the radius of gyration as a function of GdmCl activity at different temperatures (Fig. 3*A*).[‡] To allow a direct comparison with the calorimetric data, we use the same simple formalism as in ref. 45 to describe the effective binding of GdmCl to the unfolded polypeptide, assuming identical independent binding sites with an effective binding constant *K*:

$$R_g(a) = R_{g0}\left(1 + \rho\frac{Ka}{1+Ka}\right), \qquad [2]$$

where $R_{g0}$ is the radius of gyration at zero denaturant and $\rho$ is the relative change in the radius of gyration approached asymptotically at very high GdmCl activities. A van't Hoff plot of the values of *K* obtained from our data upon fitting with Eq. **2** is in surprisingly good agreement with the calorimetric data (Fig. 3*B*).[§] With the extended temperature range available here, a curvature of the plot is apparent, resulting in a heat capacity change of $\Delta c_p = 0.5 \pm 0.1$ kJ mol$^{-1}$ K$^{-1}$ and an interaction enthalpy of $\Delta h = -9 \pm 3$ kJ mol$^{-1}$ at 298 K compared with $\Delta h = -11 \pm 2$ kJ mol$^{-1}$ from the calorimetric data (assuming $\Delta c_p = 0$) (45). This agreement corroborates the intimate connection between denaturant interactions and the collapse of the unfolded state that we observe (i.e., the temperature-induced dissociation of denaturant contributes to chain compaction). However, if this were the only contribution, we would expect the curves in Fig. 3*A* to converge at zero denaturant concentration, which is not the case: The values of $R_{g0}$ obtained from the fits to Eq. **2** (Fig. 3*A*) clearly exhibit a temperature dependence (Fig. 3*C*) similar in shape to the temperature dependence at the lowest denaturant concentrations (Fig. 2*B*). We thus find a temperature-induced compaction of unfolded protein even in the absence of denaturant.

To clarify its molecular basis, we analyzed the data in Fig. 3*C* in terms of an extended form of Flory–Huggins theory (28), one of the simplest descriptions that can capture the effects of temperature on intramolecular interactions that lead to chain compaction

$$R_{g0} = \left(\frac{2nv}{3(2\chi - 1)}\right)^{1/3}, \qquad [3a]$$

where *n* is the number of monomers in the chain, *v* is the volume of a monomer, and $\chi$ is the Flory parameter, which is essentially a measure for the free energy of intrachain interactions relative to chain–solvent interactions (28). To describe the behavior of $R_{g0}(T)$, we introduce the temperature dependence of $\chi$ with

$$\chi(T) = \frac{1.4}{RT}\left(\Delta H_0 + \Delta C_p(T - T_0) - T\left(\Delta S_0 + \Delta C_p \ln\frac{T}{T_0}\right)\right) \qquad [3b]$$

where $\Delta H_0$ and $\Delta S_0$ are the enthalpic and entropic contributions, respectively, at $T_0$ and $\Delta C_p$ is the heat capacity contribution to $\chi$ (28). Interestingly, the fit to our data (Fig. 3*C*) predicts that the collapse will be followed by an increase in $R_{g0}$ at high temperature. Unfortunately, this temperature regime is currently not accessible experimentally because of the rapid loss of signal >340 K, presumably due to accelerated photodestruction of the fluorophores.

**Role of the Hydrophobic Effect.** The interaction that most obviously may be assumed to be responsible for the temperature-induced collapse of unfolded proteins is the hydrophobic effect (39, 47). Even though we do observe a significant heat capacity term for the intrachain interactions from the analysis using Eq. **3** (Fig. 3*C*), the temperature dependence does not resemble the one typically assigned to the hydrophobic effect, with its more gradual increase and maximum strength approached only at ≈400 K (47). To clarify the role of the hydrophobic effect further, we thus performed single-molecule FRET experiments on prothymosin $\alpha$, an intrin-

---

[‡] Note that for salts, such as GdmCl, the mean activity coefficient $\gamma_\pm$ of Gdm$^+$ and Cl$^-$ is used. The dimensionless activity results from the molar GdmCl concentration *c* as $a = (\gamma_\pm c\mathrm{M}^{-1})^2$. We used the tabulated values of the activity coefficient (46) for interpolation.

[§] The assumption made in Eq. **2** that the change in $R_g$ is approximately proportional to the saturation of binding sites with denaturant molecules is plausible (45), even though we are not aware of a rigorous theoretical justification. However, the values of *K* obtained from fits analogous to Eq. **2** for either $R_g$ (e.g., at 296 K, $K = 0.68 \pm 0.10$), the persistence length of the chain ($K = 0.58 \pm 0.07$), or $R_g^3$ (an effective molecular volume) ($K = 0.46 \pm 0.07$) as a function of GdmCl activity are all equal within error to the calorimetric value ($K = 0.60 \pm 0.09$ at 298 K), indicating that the choice of parameter is not particularly critical for our analysis.

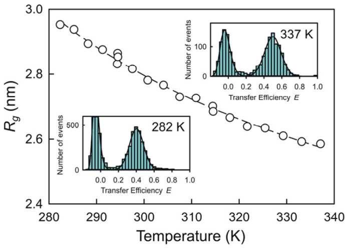

**Fig. 4.** Temperature-induced collapse of the intrinsically disordered protein prothymosin $\alpha$. The apparent $R_g$ calculated from $\langle r^2 \rangle$ of the labeled protein segment is plotted. (*Insets*) Corresponding FRET efficiency histograms at 282 and 337 K are shown as examples. An empirical fit used for interpolation is shown as a dashed line.

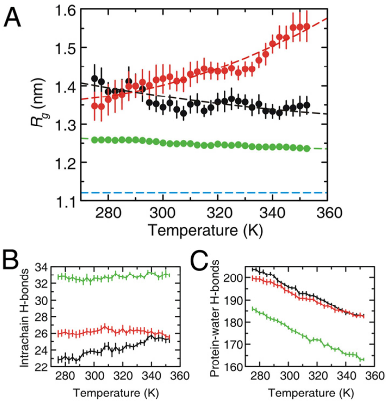

**Fig. 6.** Results on unfolded Csp*Tm* from replica exchange molecular dynamics simulations using AMBER ff03*/TIP4P-Ew water (black), AMBER ff03*/TIP3P water (red), and OPLS/AA-L/TIP3P water (green). (*A*) Radius of gyration ($R_g$) is shown as a function of temperature. Empirical fits used for interpolation are shown as dashed lines. The blue dashed line indicates $R_g$ of the folded protein. Average numbers of intramolecular hydrogen bonds (*B*) within unfolded Csp*Tm* and (*C*) between protein and water molecules as a function of temperature are shown.

sically disordered protein with extremely low hydrophobicity and an unusually large proportion of charged and polar amino acids (2, 48). In the case of a dominant contribution of the hydrophobic effect, we would expect a significantly less pronounced collapse for prothymosin $\alpha$ with increasing temperature than that for Csp*Tm*. We introduced Cys residues at positions 2 and 56 and labeled them with the same donor and acceptor fluorophores as Csp*Tm*. Single-molecule FRET experiments were performed in the presence of 0.5 M GdmCl to screen electrostatic interactions, which strongly influence the dimensions of prothymosin $\alpha$ at low ionic strengths due to its large negative net charge at pH 7 (48). Surprisingly, we observed the same degree of collapse for prothymosin $\alpha$ as for Csp*Tm*, with a reduction in $R_g$ by $\approx 13 \pm 3\%$ between 282 and 337 K (Fig. 4) compared with $13 \pm 2\%$ for Csp*Tm* in 0.5 M GdmCl. A specific effect of GdmCl can be excluded, because a collapse of the same extent is observed for prothymosin $\alpha$ in the presence of other solutes that shield the charges by an increase in ionic strength (e.g., in 0.5 M sodium phosphate) (Fig. S4). These findings suggest that the hydrophobic effect is not the main cause of temperature-induced unfolded state collapse.

An aspect that has been implicated in the behavior of unfolded proteins is the formation of secondary structure with increasing temperature. Yang et al. suggested a general propensity of unfolded polypeptides to form local extended segments at high temperatures (49), visible in NMR (50) and as an increase in the CD signal around 222 nm (ref. 49 and citations therein). To investigate the role of this effect for Csp*Tm*, we used a destabilized protein variant to exclude the influence of the unfolding transition on the observed signal change. We prepared a deletion variant lacking the five C-terminal amino acids, resulting in complete unfolding even without denaturant. To eliminate problems with aggregation at high temperatures and for direct comparability with the collapse data, the experiments were performed in 0.5 M GdmCl. A comparison of the CD spectra at 277 and 368 K (Fig. 5) shows the characteristic signal change that indicates the formation of secondary structure (49). The reversibility of the process and the absence of aggregates after return to low temperatures show that aggregation is not involved, as also observed in previous studies (49, 51). Even though the temperature dependence of the increase in secondary structure content extends over a broader temperature range than the collapse observed by FRET (Fig. 3*C*), it is suggestive of a connection between chain collapse and secondary structure content.

**Simulations.** To investigate further the microscopic origin of temperature-induced collapse, we performed replica exchange molecular dynamics simulations of unfolded Csp*Tm* in explicit water over a temperature range from 275 to 352.5 K (32 replicas) with different force fields and water models. Given that the experimental reconfiguration time of the unfolded state ranges between ≈20 and 80 ns under these conditions (26, 52) (Fig. S2), we can expect significant sampling of conformational space in the simulation time of 100–150 ns per replica (Fig. S5). Interestingly, we find that the results strongly depend on the force field and water model used. The AMBER ff03* protein force field (53) in combination with the TIP4P-Ew water model (54) resulted in a decrease of $R_g$ by $\approx 6\%$ over the entire temperature range (Fig. 6*A*), in qualitative agreement with the experimental result (Fig. 3*C*), but with TIP3P water (55), $R_g$ increased by $\approx 14\%$ (Fig. 6*A*). This difference may be related to the more accurate temperature dependence of the properties of pure water given by TIP4P-Ew compared with those given by TIP3P (54). Simulations with the OPLS-AA/L protein force field (56) also showed a slight collapse with temperature (Fig. 6*A*) but were less well converged (*SI Materials and Methods* and Fig. S6). In all cases, the mean $R_g$ in the simulations (1.24–1.56 nm) was considerably lower than that observed experimentally [≈2.2–2.4 nm, corrected for the lengths of dyes and linkers (16)] These observations stress the importance of the subtle balance between protein and water energy functions (57), which is particularly critical for simulations of unfolded proteins.

Analysis of the backbone conformations (Fig. S7*a*) indicated an increase in the population of the $\alpha$ region of the Ramachandran map; the polyproline II content decreased with temperature, similar to NMR results for a short alanine peptide (50). A DSSP (58) analysis (Fig. S7*b*), which is largely based on hydrogen bonding

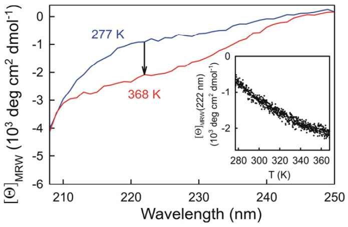

**Fig. 5.** Secondary structure content of unfolded Csp*Tm* increases with temperature. Circular dichroism spectra of Csp*Tm* destabilized by C-terminal truncation in 0.5 M guanidinium chloride at 277 K (blue) and 368 K (red). (*Inset*) Change in ellipticity at 222 nm is shown as a function of temperature.

patterns, showed an increased turn population at higher temperatures. Whereas the larger compactness in the simulations using OPLS compared with those using AMBER correlated with $\beta$-sheet content (Fig. 6*A* and Fig. S7*b*), the secondary structure changes observed with TIP3P and TIP4P-Ew were quite similar. The simulations thus may not allow us to identify a particular type of secondary structure involved in chain collapse (possibly also because longer time scales may be required to reach a completely equilibrated distribution of local secondary structure motifs), but we observed a clear overall correlation between collapse and intramolecular hydrogen bonding: in those force fields that collapsed the chain with increasing temperature, more protein–protein hydrogen bonds were formed and more protein–water hydrogen bonds lost than in the ff03*/TIP3P simulation, where the protein expanded with temperature (Fig. 6 *B* and *C*).

In summary, the simulations indicate roles for secondary structure formation and hydrogen bonding in temperature-induced collapse, but the strong dependence of the results on the water model indicate that the force fields—at least on the time scales accessible here—capture predominantly effects related to solvation.

## Discussion

The few previously reported experiments give a heterogeneous picture of the effect of temperature on unfolded state dimensions. For example, although DLS measurements on RNase A unfolded by reduction of its disulfides showed a slight compaction with temperature (37), recent SAXS measurements reported no significant change of $R_g$ (35). A significant collapse of RNase T1 in 5.3 M GdmCl was observed by DLS (38). No change in $R_g$ was reported for $\beta$-lactoglobulin in 8 M urea between 0 and 25 °C (36). Laser temperature-jump experiments on the acid-denatured small protein BBL monitored by ensemble FRET showed a clear collapse with increasing temperature (39), and NMR indicated an expansion of the cold-denatured C-terminal domain of protein L9 with decreasing temperature (59).

Here, we tried to exclude the most important complications in the investigation of unfolded state collapse. First, we used a two-state folder without disulfides to avoid ambiguities from the influence of structured intermediates or disulfides on the properties of the denatured state. Second, we used single-molecule FRET to monitor the unfolded state even under conditions where the folded state is populated, avoiding the need for additional destabilization by reduction or low pH (35, 37, 39, 40). Third, aggregation does not occur at the exceedingly low concentrations used in single-molecule experiments and can be monitored in situ (60). Fourth, we excluded a significant influence of changes in chain dynamics on the observed FRET efficiency changes by measuring nanosecond intensity correlation functions (Fig. S2), which report on the time scales of chain reconfiguration (26). Finally, we directly compared the single-molecule results with ensemble DLS experiments on the same protein without labels and under identical solution conditions to exclude a strong influence of the FRET dyes on the collapse process.

Our results reveal two conceptually different contributions to the temperature-induced collapse of unfolded proteins. One of them is the enthalpic component of the interaction between protein and denaturant, resulting in the dissociation of denaturant from the polypeptide chain with increasing temperature (45). However, denaturant dissociation cannot account for the entire effect of temperature on unfolded state collapse, as demonstrated by the change in chain dimensions at very low GdmCl concentrations (Fig. 2) and extrapolated to zero denaturant (Fig. 3*C*).[‖] Interestingly, the hydrophobic effect, the interaction most commonly associated with protein stabilization (47), cannot be the sole source of this temperature-induced compaction, because we observe a collapse of a very similar extent for two proteins with very different average hydrophobicities, Csp*Tm* and the "natively unfolded" protein prothymosin $\alpha$. Moreover, Csp*Tm* unfolds with increasing temperature, whereas the unfolded state compacts (Fig. 1), another indication that the interactions stabilizing the native state and those compacting the coil are not of the same physical origin.

Both CD measurements (Fig. 5) and molecular simulations (Fig. 6 and Fig. S7) (49) point toward a role for secondary structure formation and hydrogen bonding in temperature-induced unfolded state collapse. Even though a contribution from $\alpha$-helical conformations (23) and $\beta$ turns (61) should not be excluded, the contribution of $\beta$-structure formation may be particularly relevant. In kinetic synchrotron radiation CD experiments, we found earlier that collapsed unfolded Csp*Tm* exhibits a higher content of $\beta$ structure than the protein at high denaturant concentrations (16). Magg et al. (42) showed that the collapse of a highly homologous cold shock protein is accompanied by the local extension of a short segment that forms a $\beta$ strand in the folded structure. Most importantly, indications of an increase in $\beta$ structure with increasing temperature have been reported for several proteins (ref. 49 and citations therein), similar to the behavior that we observe for Csp*Tm* (Fig. 5). From theoretical considerations and simulations (49, 62), there has been suggested to be a general entropic bias toward extended structure, because steric interference is minimized and side chains are allowed to adopt the largest number of configurations. This effect will be amplified at high temperatures. The resulting short and probably highly dynamic $\beta$ segments may form nonnative hydrogen bonds within the unfolded polypeptide and contribute to chain compaction (19, 57). The same interactions may be responsible for the aggregation of unfolded proteins (49), a process that is particularly prevalent at high temperatures.

Relating the global compaction of unfolded proteins to local structure formation with further experiments, simulations, and theory will be interesting. Nuclear magnetic resonance, for instance, may allow an experimental determination of local structural propensities related to compaction (3, 59). By introducing systematic perturbations to explicit-solvent force fields, one may be able to identify the energetic contributions to unfolded state collapse. An alternative theoretical approach is the use of analytical models, such as structure-based free energy functionals (24) or coarse-grained simulations employing funneled energy landscapes (25). For example, in a recent study the relative contributions of native-like interactions and hydrophobic and electrostatic effects in the denatured state could be quantified (25). Such models could capture the temperature dependence of structure formation in the unfolded state if the interaction potential varied with temperature. Our results also may help to explain why many proteins exhibit two-state kinetics, since strong collapse is disfavored under folding conditions (33).

In summary, our observations bring together several independent findings that in fact may be connected closely: the temperature-induced collapse of unfolded proteins, the thermodynamics of protein–denaturant interactions, and the changes in the secondary structure propensity of unfolded polypeptides at high temperatures. The observation that the temperature dependence of attractive intramolecular interactions overcompensates for conformational entropy and leads to chain collapse with increasing temperature is in contrast to simpler systems, such as homopolymers, where the chain expands at high temperatures (28, 34). Finally, the arising picture suggests that the term "hydrophobic collapse" may be an incomplete description of the compaction of unfolded polypeptides at low denaturant concentrations and high temperatures.

---

[‖] From an analysis of the entire dataset with Eqs. **2** and **3**, we find that in >1.5 M GdmCl denaturant, dissociation accounts for ≈50% of the chain collapse between 282 and 339 K. At lower GdmCl concentrations, the denaturant-independent component dominates.

## Materials and Methods

Expression, purification, and labeling of full-length Csp*Tm* were performed as described in refs. 10 and 16. The truncated variant of Csp*Tm* and prothymosin $\alpha$ were purified using a hexahistidine tag. Single-molecule fluorescence and correlation experiments were performed using a MicroTime 200 confocal microscope (PicoQuant) essentially as described in refs. 16 and 26. Changes in refractive index and spectral properties of the dyes with GdmCl concentration and temperature were measured independently and taken into account for the calculation of $R_0$. The temperature was adjusted with a Peltier-controlled sample holder and calibrated using the temperature-dependent fluorescence lifetime of rhodamine B (63) (*SI Materials and Methods* and Fig. S1). We used a combination of additives ($\beta$-mercaptoethanol and cysteamine; see *SI Materials and Methods*) to reduce photodamage at high temperatures. Ensemble fluorescence intensity decays of the donor and acceptor were fit globally to the transfer rate distribution expected for a Gaussian chain by reconvolution with the instrument response function (16). Dynamic light scattering was measured as described in ref. 64. Circular dichroism measurements were performed at a protein concentration of 0.23 mg/mL. The absence of aggregates was tested by static light scattering in a fluorometer or by the lack of any slope in absorption measurements $>$350 nm. Replica exchange molecular dynamics simulations of unfolded Csp*Tm* were run using the OPLS/AA-L (56) and AMBER ff03* (53) force fields with two explicit water models [TIP3P (55) and TIP4P-Ew (54)], spanning a temperature range from 275 to 352.5 K in increments of 2.5 K. From unfolded configurations generated at high temperatures, simulations were run for 100–150 ns per replica (for a total of $\approx$9.3 $\mu$s). For details, see *SI Materials and Methods*.

**ACKNOWLEDGMENTS.** We thank B. Wunderlich for finite element calculations, W.A. Eaton for access to laser temperature-jump instrumentation and discussion, G. Hummer, G. Makhatadze, D. Makarov, S. Takahashi, and D. Thirumalai for valuable comments and discussions, and A. Vartapetian and A. Evstafieva for the gift of a plasmid encoding prothymosin $\alpha$. This work was supported by the Swiss National Science Foundation, the National Center of Competence in Research for Structural Biology, a Starting Researcher Grant of the European Research Council (to B.S.), and a Royal Society University Research Fellowship (to R.B.B.).

1. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
2. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
3. Mittag T, Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 17:3–14.
4. Millett IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv Protein Chem* 62:241–262.
5. Gast K, Modler AJ (2005) Studying protein folding and aggregation by laser light scattering. *Protein Folding Handbook*, eds Buchner J, Kiefhaber T (Wiley-VCH, Weinheim, Germany), pp 673–709.
6. Haran G (2003) Single-molecule fluorescence spectroscopy of biomolecular folding. *J Phys Condens Matter* 15:R1291–R1317.
7. Michalet X, Weiss S, Jäger M (2006) Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* 106:1785–1813.
8. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
9. Deniz AA, et al. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA* 97:5179–5184.
10. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
11. Magg C, Schmid FX (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein. *J Mol Biol* 335:1309–1323.
12. Laurence TA, Kong X, Jäger M, Weiss S (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc Natl Acad Sci USA* 102:17348–17353.
13. Kuzmenkina EV, Heyes CD, Nienhaus GU (2006) Single-molecule FRET study of denaturant induced unfolding of RNase H. *J Mol Biol* 357:313–324.
14. Tezuka-Kawakami T, Gell C, Brockwell DJ, Radford SE, Smith DA (2006) Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys J* 91:L42–L44.
15. Sherman E, Haran G (2006) Coil–globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
16. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
17. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci USA* 104:1528–1533.
18. Hofmann H, Golbik RP, Ott M, Hübner CG, Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* 376:597–605.
19. Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103:12394–12399.
20. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci USA* 105:13403–13408.
21. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11:83–93.
22. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
23. Wang Z, Plaxco KW, Makarov DE (2007) Influence of local and residual structures on the scaling behavior and dimensions of unfolded proteins. *Biopolymers* 86:321–328.
24. Shoemaker BA, Wolynes PG (1999) Exploring structures in protein folding funnels with free energy functionals: The denatured ensemble. *J Mol Biol* 287:657–674.
25. Weinkam P, Pletneva EV, Gray HB, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proc Natl Acad Sci USA* 106:1796–1801.
26. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
27. Cellmer T, Henry ER, Kubelka J, Hofrichter J, Eaton WA (2007) Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. *J Am Chem Soc* 129:14564–14565.
28. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 20:447–490.
29. Socci ND, Onuchic JN (1994) Folding kinetics of proteinlike heteropolymers. *J Chem Phys* 101:1519–1528.
30. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195.
31. Plotkin SS, Wang J, Wolynes PG (1996) Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top* 53:6271–6296.
32. Thirumalai D, Klimov DK (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr Opin Struct Biol* 9:197–207.
33. Chahine J, Nymeyer H, Leite VBP, Socci ND, Onuchic JN (2002) Specific and nonspecific collapse in protein folding funnels. *Phys Rev Lett* 88:168101.
34. Grosberg AY, Khokhlov AR (1994) *Statistical Physics of Macromolecules* (AIP Press, Woodbury, NY).
35. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR (2007) Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 367:609–615.
36. Katou H, Hoshino M, Kamikubo H, Batt CA, Goto Y (2001) Native-like $\beta$-hairpin retained in the cold-denatured state of bovine $\beta$-lactoglobulin. *J Mol Biol* 310:471–484.
37. Nöppert A, Gast K, Müller-Frohne M, Zirwer D, Damaschun G (1996) Reduced-denatured ribonuclease A is not in a compact state. *FEBS Lett* 380:179–182.
38. Gast K, et al. (1997) Ribonuclease T1 has different dimensions in the thermally and chemically denatured states: A dynamic light scattering study. *FEBS Lett* 403:245–248.
39. Sadqi M, Lapidus LJ, Muñoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100:12117–12122.
40. Sosnick TR, Trewhella J (1992) Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* 31:8329–8335.
41. Perl D, et al. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* 5:229–235.
42. Magg C, Kubelka J, Holtermann G, Haas E, Schmid FX (2006) Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol* 360:1067–1080.
43. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Brownian motion of ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers* 17:11–31.
44. Haas E, Wilchek M, Katchalski-Katzir E, Steinberg IZ (1975) Distribution of end-to-end distances of oligopeptides in solution as estimated by energy transfer. *Proc Natl Acad Sci USA* 72:1807–1811.
45. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226:491–505.
46. Makhatadze GI, Fernandez J, Freire E, Lilley TH, Privalov PL (1993) Thermodynamics of aqueous guanidinium hydrochloride solutions in the temperature range from 283.15 to 313.15-K. *J Chem Eng Data* 38:83–87.
47. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
48. Gast K, et al. (1995) Prothymosin $\alpha$: A biologically active protein with random coil conformation. *Biochemistry* 34:13211–13218.
49. Yang WY, Larios E, Gruebele M (2003) On the extended $\beta$-conformation propensity of polypeptides at high temperature. *J Am Chem Soc* 125:16220–16227.
50. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR (2002) Polyproline II structure in a sequence of seven alanine residues. *Proc Natl Acad Sci USA* 99:9190–9195.
51. Gast K, Zirwer D, Damaschun G (2003) Are there temperature-dependent structural transitions in the "intrinsically unstructured" protein prothymosin $\alpha$? *Eur Biophys J* 31:586–594.
52. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B* 112:6137–6146.
53. Best RB, Hummer G (2009) Optimized molecular dynamics force fields applied to the helix–coil transition of polypeptides. *J Phys Chem B* 113:9004–9015.
54. Horn HW, et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665–9678.
55. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
56. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487.
57. Bolen DW, Rose GD (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* 77:339–362.
58. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
59. Li Y, Shan B, Raleigh DP (2007) The cold denatured state is compact but expands at low temperatures: Hydrodynamic properties of the cold denatured state of the C-terminal domain of L9. *J Mol Biol* 368:256–262.
60. Hillger F, Nettels D, Dorsch S, Schuler B (2007) Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *J Fluoresc* 17:759–765.
61. Perczel A, Fasman GD (1992) Quantitative analysis of cyclic $\beta$-turn models. *Protein Sci* 1:378–395.
62. Fitzkee NC, et al. (2005) Are proteins made from a limited parts list? *Trends Biochem Sci* 30:73–80.
63. Benninger RK, et al. (2006) Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal Chem* 78:2272–2278.
64. Gast K, Nöppert A, Müller-Frohne M, Zirwer D, Damaschun G (1997) Stopped-flow dynamic light scattering as a method to monitor compaction during protein folding. *Eur Biophys J* 25:211–219.

# Supporting Information

## Nettels et al. 10.1073/pnas.0900622106

### SI Materials and Methods

**Preparation and Labeling of Proteins.** Cysteine residues were introduced by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes as described in refs. 1 and 2. Expression, purification, and labeling of full-length Csp*Tm* were performed as described in refs. 1 and 3. The truncated variant of Csp*Tm* was expressed with a cleavable hexahistidine tag to allow for rapid purification and to minimize nonspecific degradation of this unfolded protein. To this end, the gene was cloned from the vector pET21a into pET47b(+), and the sequence coding for the five C-terminal amino acids was deleted by successive quick-change mutagenesis steps. The protein was expressed in LB medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted, and DNA was digested. The cleared supernatant was loaded on a HisTrap column (GE Healthcare Bio-Sciences AB) in 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, and 4 M guanidinium chloride (GdmCl). After the 280-nm UV absorption signal reached the baseline, the column was washed with two column volumes of 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, and 10 mM imidazole, and a gradient from 10 to 500 mM imidazole was used to elute the His-tagged protein. Human rhinovirus 3C protease (containing a His tag) was added to a final concentration of 0.3 mg/mL, and after 12 h at room temperature, the cleavage reaction was dialyzed against 20 mM Tris·HCl, pH 8.0, 0.5 mM NaCl, 2 mM β-mercaptoethanol, and 10 mM imidazole and applied to a HisTrap column. The cleaved Csp*Tm* without a His tag was collected in the flow-through and concentrated. Labeling was performed as for the full-length protein.

Prothymosin α also was expressed with a cleavable hexahistidine tag for rapid purification. The coding sequence for human prothymosin α was cloned from the vector pHP12 (4) into pET47b(+). Cysteine residues were introduced in positions 2 and 56 by site-directed mutagenesis (residue numbering is from Met in the protein sequence, excluding the 19-residue N-terminal purification tag). The protein was expressed in Terrific Broth medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted, and DNA was digested. The cleared supernatant was loaded on a HisTrap column in 20 mM Tris, pH 7.0, 100 mM NaCl, and 2 mM β-mercaptoethanol. A gradient from 40 to 500 mM imidazole was used to elute the His-tagged protein. Fractions were identified via SDS/PAGE, pooled, extracted with butanol, and precipitated with ethanol (4). Pellets were dissolved in 50 mM sodium phosphate, pH 7.0, reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP), and purified on a Superdex 75 gel filtration column (GE Healthcare Bio-Sciences AB) in 100 mM sodium phosphate, pH 7.0, 2 mM β-mercaptoethanol, and 0.01% Tween. Fractions containing the full-length protein were combined, extracted with butanol, and precipitated with ethanol. The pellet was dissolved in 4 M GdmCl and 50 mM sodium phosphate, pH 7.0, and the protein concentration was determined by bicinchoninic acid assay (BCA Protein Assay Kit; Pierce). Fluorophore labeling was performed at a protein concentration of ≈0.1 mg/mL with a threefold excess of Alexa Fluor 488 and Alexa Fluor 594; gel filtration was used to remove the free dye. The correct molecular mass of the labeled protein was confirmed by MALDI mass spectrometry. The donor-only labeled and acceptor-only labeled prothymosin α resulting from such random labeling does not interfere with single-molecule FRET measurements.

**Single-Molecule Fluorescence Spectroscopy.** Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) equipped with a continuous-wave 488-nm diode laser (Sapphire 488–100 CDRH; Coherent) and an Olympus UplanApo 60×/1.20W objective. Sample fluorescence was separated into donor and acceptor components using a dichroic mirror (585DCXR; Chroma Technology) and two final filters (ET525/50M and HQ650/100; Chroma Technology). Each component was focused onto an avalanche photodiode (SPCM-AQR-15; PerkinElmer Optoelectronics), and the arrival time of every detected photon was recorded. Samples of labeled protein were diluted to a concentration of ≈20 pM in 50 mM sodium phosphate buffer at the appropriate GdmCl (Pierce) concentration, individually adjusted to pH 7. Tween 20 (0.001%; Pierce) was added to prevent surface adhesion of the protein (1). To reduce data acquisition time and thus minimize damage to the chromophores at high temperatures, the photoprotective additives β-mercaptoethanol (200 mM) and cysteamine (5 mM) were included. As a result, the measurements could be performed at a laser power of 400 μW at the sample, resulting in significantly more intense fluorescence bursts and a reduction of the acquisition time from approximately 1 h to 15 min per measurement (for 5,000–10,000 identified bursts). Successive photons detected in either channel separated by <100 μs were combined into one burst. Identified bursts were corrected for background, differences in quantum yields of the donor and acceptor, different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation, as described in ref. 5. A burst was retained as a significant event if the total number of counts exceeded 70.

The Förster radius $R_0$ was corrected for the changes in solution conditions, which were dominated by the change in refractive index $n$ with GdmCl concentration (2). The overlap integral (6) for Alexa Fluor 488 emission and Alexa Fluor 594 absorption was found to be independent of GdmCl concentration and temperature within error of the measurement. The change in $n$ with temperature was measured refractometrically; it was found to be <1% over the temperature range used here and very similar to the refractive index change of water with temperature (7) at all GdmCl concentrations. Thus, $n$ is dominated by the change in GdmCl concentration. Assuming the change of $n$ with absolute temperature $T$ and GdmCl concentration $c_\mathrm{D}$ (8) to be mutually independent, $n(T, c_\mathrm{D})$ was approximated by the following interpolation function

$$\begin{aligned} n(T, c_D) = {} & 1.3338 + 0.0174716\ \mathrm{M}^{-1}\, c_D - 0.0001855\ \mathrm{M}^{-2}\, c_D^2 \\ & + 7.80416 \times 10^{-6}\ \mathrm{M}^{-3} c_D^3 + 0.000262 \\ & (323.5 - 48.5\ \mathrm{e}^{0.02062(275 - T/K)} - T/K). \end{aligned}$$

Fluorescence lifetime experiments on Csp*Tm* singly labeled with the donor or acceptor show a negligible change of the fluorescence lifetime (and thus the quantum yield) with temperature. Correlation experiments (9) show that under the conditions used here chain reconfiguration times are at least approximately an order of magnitude greater than the donor fluorescence lifetime (Fig. S2) but small relative to the mean burst duration of ≈1 ms, justifying the use of Eq. **1** (main text) (10), which neglects the influence of chain diffusion on the observed transfer efficiencies (11). By numerically solving the diffusion equation in the potential of mean force corresponding to $P(r)$ for a Gaussian chain modified by a sink term to account for FRET (11), with the intramolecular end-to-end diffusion coefficients obtained from correlation experiments (9) as a function of temperature (Fig. S2), we can estimate the effect of the change in intramolecular diffusion on the change in the radius of gyration ($R_{\mathrm{g}}$) over the accessible temperature range to be $\approx$10% of the observed amplitude (i.e., less than the experimental uncertainty). Time-resolved fluorescence polarization anisotropy decays indicate rapid reorientation of the dyes during the fluorescence lifetime of the donor (12); an orientational factor $\kappa^2$ (6) of 2/3 thus can be used to calculate $R_0$ (2). The transfer efficiency of the unfolded state was independent of the laser power in the range of excitation rates used here.

A custom-built temperature-controlled sample holder employing Peltier elements and a digital temperature controller (TC2812-LAB12; Cooltronic) with a PT100 temperature sensor was used (Fig. S1). Because of the danger of damage to the microscope objective at high temperatures, the objective cannot be heated to the highest temperatures required here. The contact of the sample cell to the objective via immersion water thus results in a temperature gradient in the sample, and the temperature needs to be determined directly at the confocal volume. It was calibrated via the temperature-dependent fluorescence lifetime of rhodamine B (13) determined in a custom-built temperature-controlled ensemble time-correlated single photon counting instrument (12), where the temperature can be measured with a thermocouple directly in the sample whose temperature is kept uniform by stirring. Rhodamine B lifetimes from time-correlated single photon counting measurements in the confocal instrument under identical conditions then can be converted to temperature with an accuracy of approximately ±2 K. The sample was excited at 470 nm with a 20-MHz repetition rate using a pulsed diode laser (LDH 470; PicoQuant); emission was observed at magic-angle settings using a z 582/15 filter (Chroma Technology). For measurements on labeled Csp*Tm* in 5.3 M GdmCl, ET525/50M and HQ650/100 filters (Chroma Technology) were used for the donor and acceptor emission, respectively. The decays were fit globally to the transfer rate distribution expected for a Gaussian chain by reconvolution with the instrument response function (FWHM of 80 ps) (2). Note that at the extremely low excitation rate of these ensemble experiments, the donor-only peak is absent in our Csp*Tm* preparation, as in single-molecule experiments at very low power (14) or measurements under flow (14), and therefore does not influence the analysis.

**Dynamic Light Scattering.** Dynamic light scattering was measured at a scattering angle of 90° and a wavelength of 532 nm using a custom-built apparatus, equipped with a diode-pumped, continuous-wave laser (Millennia IIs; Spectra-Physics) and an avalanche photodiode. Details of the detection and data processing procedures have been described elsewhere (15). The translational diffusion coefficients $D$ were obtained from the measured autocorrelation functions using the program CONTIN (16). The translational diffusion coefficients $D$ were converted to hydrodynamic radii via the Stokes–Einstein equation $R_{\mathrm{h}} = k_{\mathrm{B}}T/(6\pi\eta_0 D)$, where $k_{\mathrm{B}}$ is Boltzmann's constant, $T$ is absolute temperature, and $\eta_0$ is the solvent viscosity. Solvent viscosities and densities were measured using an Ubbelohde type viscometer (Viscoboy 2; Lauda) and a digital density meter (DMA 58; Anton Paar), respectively. Refractive indices of the solvent mixtures were measured with an Abbe type refractometer. The solvents and protein solutions were filtered through 100-nm pore-size Anotop filters (Whatman) directly into small flow cells with path lengths of 1.5 mm (Hellma).

**Circular Dichroism Spectroscopy.** Experiments were performed in a Jasco J-715 CD spectrometer (Jasco) equipped with a temperature-controlled water bath, using a cylindrical quartz cell with a path length of 1 mm and a protein concentration of 0.23 mg/mL in 50 mM sodium phosphate buffer with 0.5 M GdmCl adjusted to pH 7.0. The absence of aggregates was tested by static light scattering in a fluorometer or by the lack of any slope in absorption measurements at wavelengths $>$350 nm.

**Molecular Dynamics Simulations.** Simulations of unfolded Csp*Tm* were run in explicit solvent using periodic boundary conditions with the GROMACS 4.0.3 package (17). Three sets of simulations were run: (*i*) with the AMBER ff03* force field (18) for the protein with TIP4P-Ew water (19) in a truncated octahedral cell with an initial size of 65 Å, containing 6,637 water molecules; (*ii*) with the AMBER ff03* force field (18) for the protein with TIP3P water (20) in a truncated octahedral cell with an initial size of 65 Å, containing 6,637 water molecules; (*iii*) with the OPLS/AA-L force field (18) for the protein with TIP3P water (20) in a cubic cell with an initial size of 60 Å, containing 6,648 water molecules. The simulations were propagated using Langevin dynamics with a time step of 2 fs and a friction coefficient of 1 ps$^{-1}$ on all heavy atoms to maintain a constant temperature, and bond lengths were kept fixed using the LINCS constraint algorithm (21). Long-range electrostatics were calculated using particle mesh ewald (PME) with a grid spacing of 1.2 Å and a cutoff of 9 Å. Initial configurations were obtained by solvating an unfolded chain in a box of water and deleting the overlapping water molecules. To obtain the correct size for the simulation cell, the system was equilibrated for 200 ps at 300 K at a constant pressure of 1 atm using a Parrinello–Rahman barostat (22). From the equilibrated configuration, a 1-ns simulation at constant volume was run at 800 K to generate randomized initial coordinates. Replica exchange molecular dynamics at a constant volume was performed using 32 replicas spanning the temperature range from 275 to 352.5 K in increments of 2.5 K, with each replica initiated from a different configuration drawn from the run at 800 K. Exchange attempts were made every 5 ps, and coordinates were saved every 10 ps. Each replica was run for 140, 150, and 100 ns for simulations *i*, *ii*, and *iii*, respectively. Fig. S5 shows the cumulative $R_{\mathrm{g}}$ over the simulations, indicating that most of the relaxation to equilibrium is complete after $\approx$30 ns. The first 30 ns of the simulations thus were discarded to allow for equilibration. A comparison of the radii of gyration and end-to-end distances as a function of temperature is given for all three force fields in Fig. S6. The data for simulation *iii* appear to be less well converged than those for the other two force fields, as suggested by the increase in end-to-end distance for this simulation, while the overall $R_{\mathrm{g}}$ decreases (Fig. S6).

1. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
2. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
3. Kremer W, et al. (2001) Solution NMR structure of the cold-shock protein from the hyperthermophilic bacterium *Thermotoga maritima*. *Eur J Biochem* 268:2527–2539.
4. Evstafieva AG, et al. (1995) Overproduction in *Escherichia coli*, purification and properties of human prothymosin $\alpha$. *Eur J Biochem* 231:639–643.
5. Schuler B (2007) Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol Biol* 350:115–138.
6. Van Der Meer BW, Coker G, III, Chen, S-YS (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).
7. Abbate G, Bernini U, Ragozzino E, Somma F (1978) The temperature dependence of the refractive index of water. *J Phys D Appl Phys* 11:1167–1172.
8. Nozaki Y (1972) The preparation of guanidine hydrochloride. *Methods Enzymol* 26:43–50.
9. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
10. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci USA* 102:2754–2759.

11. Haas E, Katchalski-Katzir E, Steinberg IZ (1978) Brownian-motion of ends of oligopeptide chains in solution as estimated by energy transfer between the chain ends. *Biopolymers* 17:11–31.
12. Nettels D, Hoffmann A, Schuler B (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. *J Phys Chem B* 112:6137–6146.
13. Benninger RK, et al. (2006) Quantitative 3D mapping of fluidic temperatures within microchannel networks using fluorescence lifetime imaging. *Anal Chem* 78:2272–2278.
14. Lipman EA, Schuler B, Bakajin O, Eaton WA (2003) Single-molecule measurement of protein folding kinetics. *Science* 301:1233–1235.
15. Gast K, Nöppert A, Müller-Frohne M, Zirwer D, Damaschun G (1997) Stopped-flow dynamic light scattering as a method to monitor compaction during protein folding. *Eur Biophys J* 25:211–219.
16. Provencher SW (1982) CONTIN: A general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput Phys Commun* 27:229–242.
17. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
18. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105:6474–6487.
19. Horn HW, et al. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* 120:9665–9678.
20. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
21. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472.
22. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52:7182–7190.

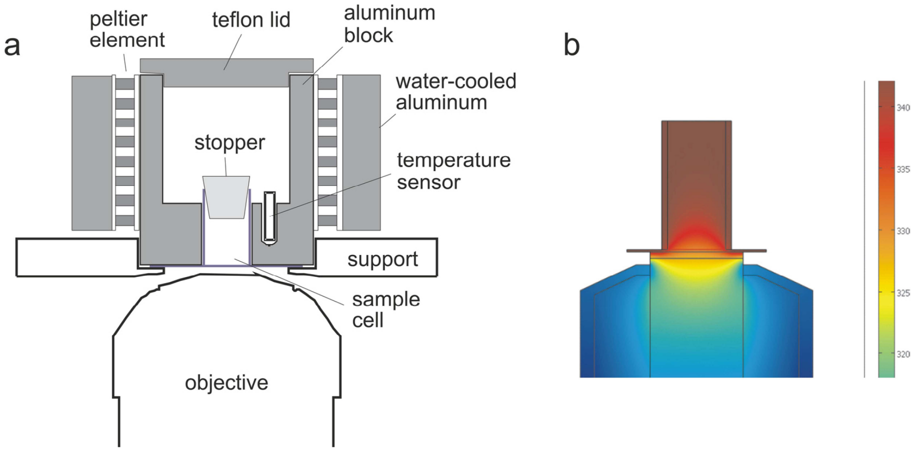

**Fig. S1.** Temperature-controlled sample holder for confocal single-molecule measurements. (*a*) Schematic of the Peltier-controlled sample holder. Evaporation of immersion water is avoided by a rubber sealing ring between the objective and the cell support. (*b*) The temperature distribution from finite element calculations (objective at room temperature and sample holder at 342 K, prepared using COMSOL Multiphysics; COMSOL) shows a pronounced temperature gradient, illustrating that the temperature must be determined locally (i.e., at the position of the confocal volume just above the cover slide of the sample cell). The calculated temperature at the position of the confocal volume is in good agreement with the values obtained from rhodamine B fluorescence lifetimes.

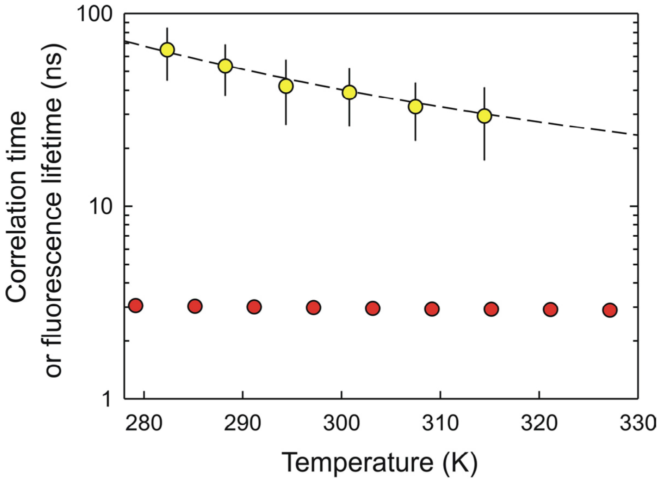

**Fig. S2.** Temperature dependence of the decay time of the donor fluorescence intensity correlation function (yellow), measured as described in ref. 9, and of the average fluorescence lifetime of the donor (red) of FRET-labeled full-length Csp*Tm* (C2C67) in 5.3 M guanidinium chloride. The dashed line shows a fit to the data assuming that the correlation times scale with the temperature dependence of the viscosity of water. The good agreement indicates that the decrease in solvent viscosity is the dominant cause of faster chain dynamics with increasing temperatures. The error bars indicate our estimate of the experimental uncertainty.

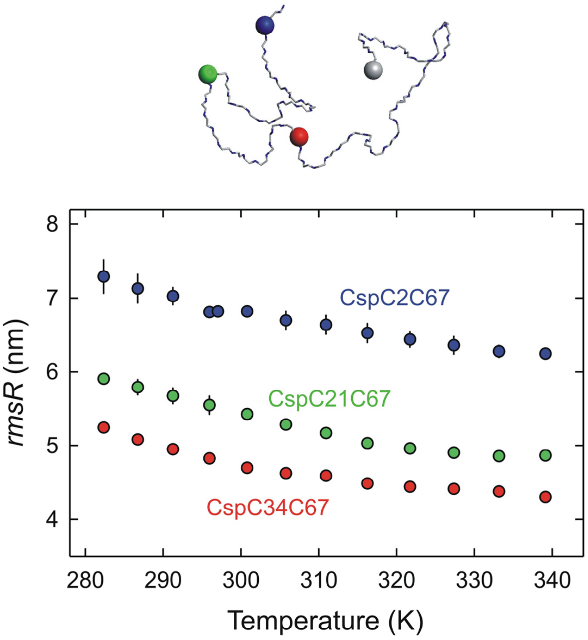

**Fig. S3.** Temperature-induced unfolded state collapse of Csp*Tm* variants with the FRET pair in different positions within the polypeptide chain in 2.0 M guanidinium chloride. Data acquisition and analysis were performed as in Figs. 1 and 2. Fluorophores were placed at positions 2 and 67 (blue, same variant as in Figs. 1, 2, and 3), 21 and 67 (green), or 34 and 67 (red, illustrated in a schematic of the unfolded chain above the graph). Variants were prepared as described in ref. 2. Error bars indicate standard deviations estimated from at least two independent measurements. The similarity of the changes in intramolecular root-mean-squared distances (*rmsR*) of 14 ± 3% (CspC2C67), 17 ± 2% (CspC21C67), and 18 ± 2% (CspC34C67) provides additional evidence that the temperature-induced compaction of unfolded Csp*Tm* is a global process that affects the entire chain.

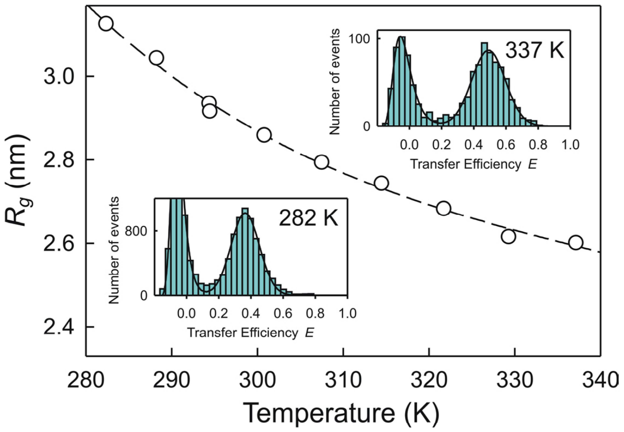

**Fig. S4.** Temperature-induced collapse of prothymosin $\alpha$ in 0.5 M sodium phosphate buffer. (*Insets*) Corresponding FRET efficiency histograms at 282 and 337 K as examples. An empirical fit used for interpolation is shown as a dashed line.

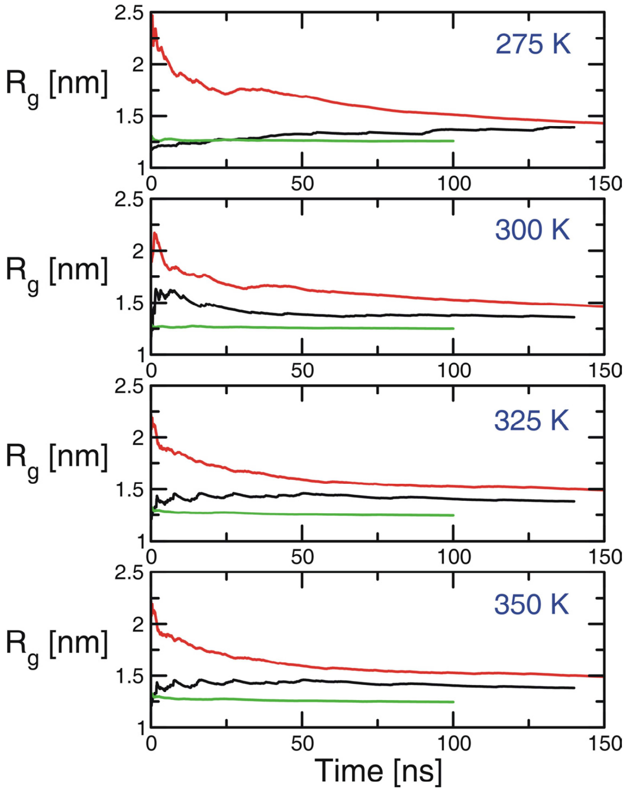

**Fig. S5.** Cumulative average radius of gyration ($R_g$) of unfolded Csp*Tm* as a function of temperature for the three sets of simulations: (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green).

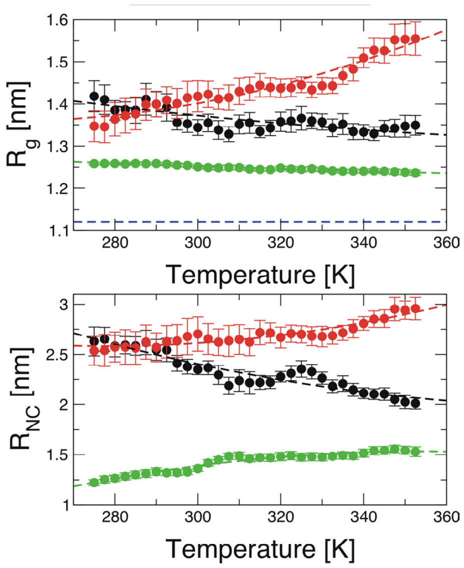

**Fig. S6.** Radius of gyration ($R_g$) of unfolded Csp*Tm* (*upper*) as a function of temperature for the three sets of simulations, (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green), and the corresponding mean end-to-end distance ($R_{NC}$) (*lower*).

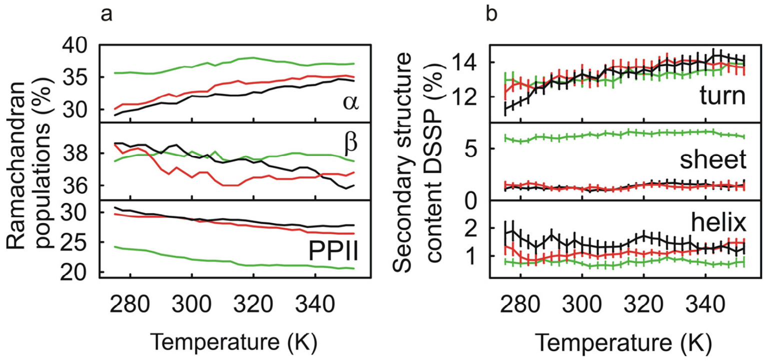

**Fig. S7.** Populations in a Ramachandran map (*a*) as a function of temperature for the three sets of simulations, (*i*) AMBER ff03*, TIP4P-Ew water (black), (*ii*) AMBER ff03*, TIP3P water (red), and (*iii*) OPLS/AA-L, TIP3P water (green), and average secondary structure content from a Database of Secondary Structure of Proteins (DSSP)-based analysis (*b*).

# 2. Charge interactions can dominate the dimensions of intrinsically disordered proteins

*Proc Natl Acad Sci U S A*, **2010**, published online, doi: 10.1073/pnas.1001743107

# Charge interactions can dominate the dimensions of intrinsically disordered proteins

**Sonja Müller-Späth[1], Andrea Soranno[1], Verena Hirschfeld[3], Hagen Hofmann, Stefan Rüegger[4], Luc Reymond[5], Daniel Nettels, and Benjamin Schuler[2]**

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Edited by Devarajan Thirumalai, University of Maryland, College Park, MD, and accepted by the Editorial Board June 9, 2010 (received for review February 18, 2010)

**Many eukaryotic proteins are disordered under physiological conditions, and fold into ordered structures only on binding to their cellular targets. Such intrinsically disordered proteins (IDPs) often contain a large fraction of charged amino acids. Here, we use single-molecule Förster resonance energy transfer to investigate the influence of charged residues on the dimensions of unfolded and intrinsically disordered proteins. We find that, in contrast to the compact unfolded conformations that have been observed for many proteins at low denaturant concentration, IDPs can exhibit a prominent expansion at low ionic strength that correlates with their net charge. Charge-balanced polypeptides, however, can exhibit an additional collapse at low ionic strength, as predicted by polyampholyte theory from the attraction between opposite charges in the chain. The pronounced effect of charges on the dimensions of unfolded proteins has important implications for the cellular functions of IDPs.**

polyampholyte | polyelectrolyte | protein folding | unfolded state | single-molecule FRET

A surprisingly large number of proteins contain extended unstructured segments or fold into a well-defined three-dimensional structure only in the presence of their specific ligands or binding partners (1–4). Especially in eukaryotes, such intrinsically disordered or unstructured proteins (IDPs) appear to be involved in a wide range of cellular functions, including transcription, translation, signal transduction, and the regulation of protein assembly (1). Correspondingly, many IDPs are associated with diseases, such as cancer or neurodegenerative disorders (4). In contrast to stably folded proteins, the polymer properties of IDPs are crucial for many of their functions. Long-range interactions of the unstructured chain with binding partners have been suggested to increase their capture radius, leading to an enhancement of binding rates via the "fly-casting" mechanism (5, 6). Conformational disorder has an important role in mediating binding diversity, enabling interactions with multiple targets (2, 7, 8). Repulsive entropic forces of "brush-like" structures can give rise to very long-range interactions, for instance to provide a mechanism for maintaining neurofilament and microtubule spacing (9, 10). Finally, because part of the binding free energy has to be expended for folding of the IDP upon interaction with its target, intrinsic disorder facilitates highly specific binding at moderate affinities, a mechanism that may be essential for regulation and signal transduction (2). All of these aspects depend crucially on chain flexibility and dimensions, which in turn are dictated by the composition of the polypeptide. Establishing a quantitative relation between charge content, hydrophobicity, and chain dimensions is thus a prerequisite for understanding the molecular mechanisms underlying IDP function.

Prevalent characteristics of IDPs are their low sequence complexity, the low proportion of hydrophobic residues, which usually form the core of a folded protein, and their high content of polar and charged amino acids (11). This correlation between disorder and charge content has strongly aided the identification of IDPs from large scale genomic sequence data bases and now provides an ideal opportunity to investigate the role of sequence composition and especially charge interactions for the conformational properties of unfolded proteins. It also allows us to test quantitative descriptions and predictions of polymer theory for the influence of charged amino acids on chain dimensions. Even though the use of the corresponding simplified models would seem ideally suited for developing a better understanding of the properties of IDPs, there have been few, if any, measurements available that would have allowed a quantitative comparison to polymer theory (9).

Here, we use single-molecule Förster resonance energy transfer (FRET), a method that has been applied successfully to obtain long-range distance distributions and dynamics in unfolded proteins (12–14). By means of the spectroscopic separation of folded and unfolded subpopulations (15), single-molecule FRET allows us to distinguish changes in the conformational properties within one of the subpopulations from a change in their relative abundances, which is often difficult in corresponding ensemble experiments. To sample the range of sequence compositions found in natural proteins, as represented by a plot of net charge versus hydrophobicity calculated from the amino acid sequence (11) (Fig. 1), we chose three representative proteins with very different properties: the globular cold shock protein Csp*Tm*, which is stably folded even in the absence of ligands (16–18); the N-terminal domain of HIV-1 integrase (IN), which folds only upon binding of $Zn^{2+}$ ions and is otherwise denatured (19, 20); and human prothymosin α (ProTα), one of the IDPs with the largest fraction of charged amino acids identified so far (11). ProTα does not assume a well-defined folded structure under any known conditions and does not contain regular secondary structure (21), but plays crucial roles in different biological processes including cell proliferation, transcriptional regulation, and apoptosis (22).

## Results

**Single-Molecule FRET Experiments.** We labeled all protein variants with a donor (Alexa Fluor 488) and an acceptor (Alexa Fluor 594) fluorophore (*Methods* and *SI Text*), and investigated them with confocal single-molecule fluorescence spectroscopy. For IN and Csp*Tm*, the labels were positioned close to the termini of the polypeptides (Fig. 1), such that the properties of the entire

---

Author contributions: S.M.-S., A.S., and B.S. designed research; S.M.-S., A.S., and V.H. performed research; S.M.-S., A.S., H.H., S.R., L.R., and D.N. contributed new reagents/analytic tools; S.M.-S., A.S., and H.H. analyzed data; and S.M.-S., A.S., and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.T. is a guest editor invited by the Editorial Board.

[1] S.M.-S. and A.S. contributed equally to this work.

[2] To whom correspondence should be addressed. E-mail: schuler@bioc.uzh.ch.

[3] Present address: Institut für Physik, Universität zu Lübeck, 23538 Lübeck, Germany.

[4] Present address: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.

[5] Present address: Laboratoire d'ingénierie des protéines, EPFL, 1015 Lausanne, Switzerland.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1001743107/-/DCSupplemental.

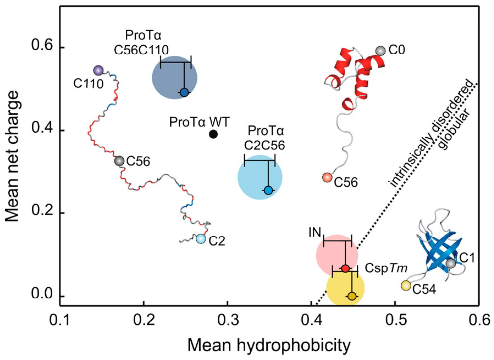

**Fig. 1.** Mean net charge versus mean hydrophobicity per residue of the globular and intrinsically disordered proteins used in this study. The dotted line indicates the separation between intrinsically disordered and globular proteins observed by Uversky et al. (11). Small circles are based on calculations taking into account the amino acid sequence only. Vertical bars indicate the influence of the dye charges; horizontal bars are an estimate of the uncertainty in the hydrophobicity of the dyes. To estimate this uncertainty, the hydrophobicity of the residues at the position of fluorophore attachment was varied between the value for the most hydrophilic and the most hydrophobic amino acid, and the resulting average hydrophobicities were computed. The large circles illustrate this range of values. For Csp*Tm* and IN, where the dyes were positioned close to the termini, the entire sequence was used for calculating charge and hydrophobicity, for the ProTα variants only the interdye segment. Hydrophobicity values were calculated according to Kyte and Doolittle (67). The positions used for FRET labeling are indicated as small spheres in the structural representations of the proteins.

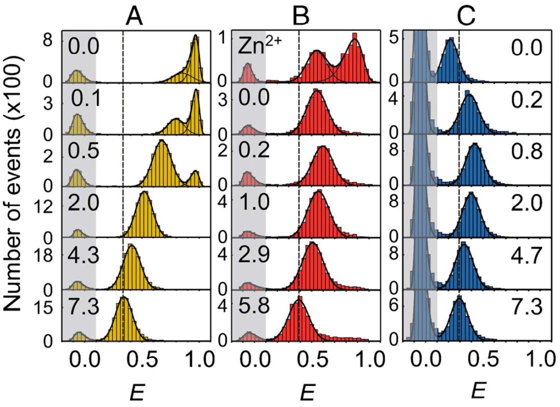

**Fig. 2.** Single-molecule FRET efficiency ($E$) histograms of (*A*) Csp*Tm* (C-terminally truncated variant), (*B*) IN, and (*C*) ProTαC show the GdmCl dependence of the unfolded proteins. The molar GdmCl concentration is indicated in each panel. (*A*) The peak at $E \approx 0.95$ corresponds to folded Csp*Tm*, the peak between $E \approx 0.3$ and 0.85 to unfolded protein. (*B*) First panel: Folded IN ($E \approx 0.9$) is only populated in the presence of $ZnCl_2$ (100 μM; 0 M GdmCl). Other panels: varying concentrations of GdmCl with 1 mM EDTA. $\langle E \rangle$ of the unfolded population ranges between 0.4 and 0.6. (*C*) ProTαC is unfolded under all conditions. For all proteins, the peaks at $E \approx 0$ (shaded) correspond to molecules lacking an active acceptor chromophore (68). The solid lines show fits used to extract the mean transfer efficiencies of the subpopulations (17, 23). The dashed lines indicate the mean transfer efficiencies of the unfolded states at the highest GdmCl concentrations.

chain were probed. For ProTα, the N- and C-terminal segments of the polypeptide exhibit very different charge densities and were investigated separately by positioning one chromophore at position 56 and the other either at position 2 (ProTαN) or at position 110 (ProTαC; see Table S1). In this way, the length of all segments probed was similar and resulted in average interdye-distances sufficiently close to the Förster radius of the dye pair to optimize sensitivity and simplify a quantitative comparison. The efficiency of energy transfer between the dyes upon donor excitation was determined from photon bursts originating from individual molecules freely diffusing through the focal spot of the laser beam as $E = n_A/(n_A + n_D)$, where $n_A$ and $n_D$ are the number of detected acceptor and donor photons, respectively [including corrections (*SI Text*)]. A transfer efficiency histogram generated from a large number of such events shows distinct maxima corresponding to the subpopulations present in the sample (Fig. 2).

Fig. 2 shows examples of FRET efficiency histograms obtained under different solution conditions. For Csp*Tm* (Fig. 2*A*), we used a variant destabilized via a C-terminal truncation of two residues. As a result, the unfolded state is populated in the absence of denaturant, which allows us to determine the unfolded state dimensions of Csp*Tm* even under these conditions. The peak at high $E$ corresponds to folded molecules, the peak at intermediate $E$ corresponds to unfolded molecules, and the peak at $E \approx 0$ (shaded) originates from molecules with an inactive acceptor. With increasing guanidinium chloride (GdmCl) concentration, the population of folded molecules decreases, and the population of unfolded molecules increases, as expected for a two-state system (15, 23). At the same time, the mean transfer efficiency $\langle E \rangle$ of the unfolded state decreases continuously (Fig. 2*A*), corresponding to the well-studied denaturant-induced expansion of unfolded Csp*Tm* (17, 23, 24), a behavior that has now been observed for the unfolded states of many proteins (12). With the ability to separate subpopulations, single-molecule experiments allow such continuous changes in the dimensions of the unfolded state to be clearly distinguished from the contribution of the native state or other populations to the overall signal.

In contrast to Csp*Tm*, IN (Fig. 2*B*) is completely unfolded even in the absence of denaturant; i.e., it is in its intrinsically disordered state. Only in the presence of $ZnCl_2$, a folded subpopulation of molecules coexists with unfolded IN. On addition of EDTA, which complexes $Zn^{2+}$ with high affinity, all molecules unfold. Interestingly, the denaturant dependence of unfolded IN shows a nonmonotonic behavior: starting from 0 M GdmCl, $\langle E \rangle$ for unfolded IN first increases slightly up to GdmCl concentrations of about 0.2 M, indicating a collapse of the unfolded state. Only at higher GdmCl concentrations, $\langle E \rangle$ starts to decrease due to the denaturant-induced expansion. This effect is even more pronounced for ProTα (Fig. 2*C*, variant ProTαC), with a drastic increase in $\langle E \rangle$ from 0.21 in the absence of GdmCl to 0.43 at 0.8 M GdmCl, followed by the denaturant-induced expansion similar to Csp*Tm* and IN at higher GdmCl concentrations. At 0 M denaturant, ProTαC is thus more expanded than at the highest accessible GdmCl concentrations.

The denaturant dependences of the three proteins are summarized in Fig. 3. It shows the monotonic change in $\langle E \rangle$ for unfolded Csp*Tm*, corresponding to its continuous expansion with increasing GdmCl concentration, and for the IDPs the remarkable "rollover" of $\langle E \rangle$ below approximately 0.5 M GdmCl. The correlation between the amplitude of the rollover and the charge density of the protein strongly suggests that electrostatic repulsion within the polypeptides cause the expansion of the IDPs. In the presence of the ionic denaturant GdmCl, the charges of the amino acid side chains are screened, allowing the polypeptides to compact (25)*. Only at higher concentrations of GdmCl, the denaturant-induced expansion of the chain takes over, and ultimately the transfer efficiencies of the different unfolded proteins converge (Fig. 3), as expected for polypeptides of similar length (26)[†]. The rollover is absent if the uncharged

---

*A similar compaction can be achieved by adding other salts (see below).

[†] The slightly lower transfer efficiency of ProTα at high GdmCl concentrations compared to the other proteins is presumably an excluded volume effect of the large unlabeled segment present in the ProTα chain.

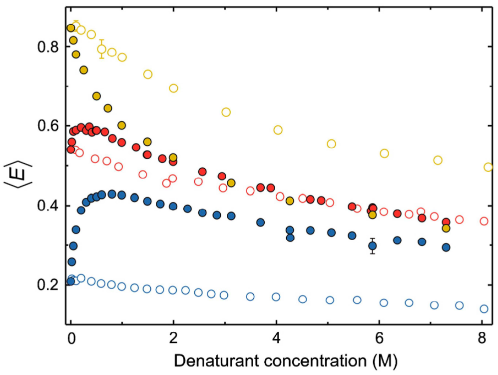

**Fig. 3.** Denaturant-dependent collapse and charge-mediated expansion of unfolded proteins. Dependence of the mean transfer efficiencies, ⟨*E*⟩, for Csp*Tm* (yellow), IN (red), and ProTαC (blue) on the concentration of GdmCl (filled circles) and urea (open circles). The typical uncertainty in transfer efficiency of individual data points is in the range of 0.02. Error bars are shown for conditions where multiple measurements are available.

denaturant urea is used. Moreover, the differences in transfer efficiencies between the proteins in the absence of denaturant are present over the entire range of urea concentrations, indicating that charge-mediated repulsion dominates chain dimensions even at the highest concentrations of urea.

To facilitate a quantitative analysis of our observations in terms of unfolded state dimensions, we converted the mean transfer efficiencies to a measure of intramolecular distance according to (23)

$$\langle E\rangle = \int_0^\infty E(r)P(r)dr \quad \text{with } E(r) = 1/(1+(r/R_0)^6), \qquad [\mathbf{1}]$$

where $r$ is the distance between donor and acceptor, $P(r)$ is the normalized equilibrium distance distribution, and $R_0$ is the Förster radius (5.4 nm at 0 M denaturant) calculated for the respective solution conditions. Given a measured value of $\langle E\rangle$ of the unfolded subpopulation and a suitable model for the distance distribution, the parameters determining $P(r)$ can be calculated numerically. To test the robustness of our results with respect to the functional form of the distance distribution used, we analyzed the data with two different models: the Gaussian chain (12, 17, 23, 27, 28) and a variation of Sanchez theory (28–30), which uses a Flory–Fisk distribution (31) with a solvent-dependent effective interaction within the chain (see *SI Text* for details). The results from the two types of analysis are very similar (Fig. S1, Fig. S2, Table S2, Table S3, Table S4, and Table S5), suggesting that the chain dimensions derived from the FRET data under our conditions do not strongly depend on the assumptions underlying the individual models. We thus proceed with the dimensions calculated using the simpler model, the Gaussian chain[‡].

Fig. 4 shows the resulting values of the radius of gyration, $R_g$ (Eq. **S1**), as a function of denaturant concentration for all protein variants investigated. Whereas their behavior is similar at GdmCl concentrations above approximately 1 M, the extent of their charge-driven unfolded state expansion at low GdmCl concentration is very different, increasing in amplitude from 8% of the $R_g$ in IN (Fig. 4*B*) and 14% in ProTαN (Fig. 4*C*) to 46% in ProTαC (Fig. 4*D*). The degree of expansion clearly correlates with the net charge of the proteins: Csp*Tm* is almost charge-balanced (net charge −2)[§]; IN shows a slight excess of negative charges (net charge −4), and the two ProTα segments ProTαN (net charge −14) and ProTαC (net charge −27) exhibit a large net charge. This correlation suggests an important role of charge density in the polypeptides for their unfolded state dimensions. To test this hypothesis and to identify a suitable quantitative description of IDPs (32), we applied a polymer physical analysis.

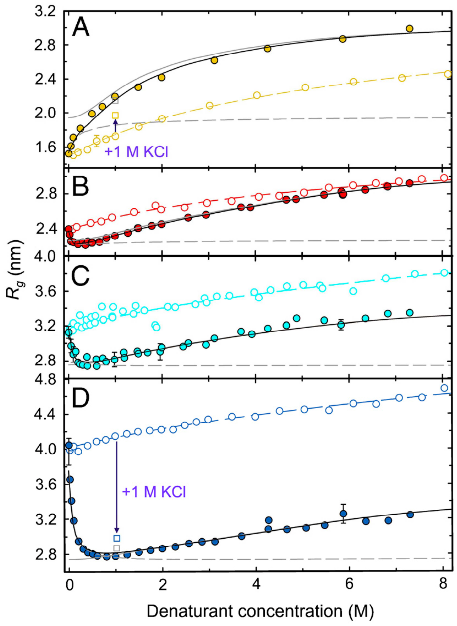

**Fig. 4.** Dependence of the apparent radii of gyration ($R_g$) of the labeled protein segments on the concentration of GdmCl (filled circles) and urea (open circles), with (*A*) Csp*Tm* (yellow), (*B*) IN (red), (*C*) ProTαN (cyan), and (*D*) ProTαC (blue). Fits to a binding model for the urea dependence (Eq. **2**, colored dashed lines), and to polyampholyte theory for the GdmCl dependence (Eq. **5**, black solid lines) are shown. The two components of Eq. **5**, corresponding to the contributions of GdmCl binding and electrostatic repulsion, are indicated as continuous and dashed gray lines, respectively. Note that the fits to Eq. **5** are performed based on thermodynamic activities, but plotted on a concentration scale. The colored squares in (*A*) and (*D*) indicate the values of $R_g$ on addition of 1 M KCl (compare to Fig. 5). The gray squares indicate the expected values estimated with Eqs. **4** and **5**, assuming the values for $K$, $a$, and $\rho$ obtained from the fits of the urea dependencies (Table S4), the value of $\nu$ obtained from the fits of the GdmCl dependencies (Table S2), and calculating $\kappa$ for an ionic strength of 1 M. The remaining difference between experimental and calculated values may be due to the preferential interaction of GdmCl with the polypeptide, leading to a higher local charge density than in the bulk solution and a correspondingly stronger charge shielding than for KCl.

**Quantifying Charge Effects on Unfolded State Dimensions.** At least three contributions need to be taken into account to describe our results: (*i*) the expansion of the polypeptide with increasing denaturant concentration, (*ii*) the electrostatic interactions between the charges in the chain, and (*iii*) the screening of these charges by the ionic denaturant GdmCl. To describe chain expansion with increasing denaturant concentration, we use a simple binding model (17, 18, 33) that assumes identical independent binding sites with an effective association constant $K$

$$R_g(a) = R_{g0}\left(1+\rho\frac{Ka}{1+Ka}\right), \qquad [\mathbf{2}]$$

where $R_{g0}$ is the radius of gyration at zero denaturant, $a$ is the thermodynamic activity of denaturant (18, 33), and $\rho$ accounts for the relative change in radius of gyration approached asymptoti-

---

[‡] Note also that unfolded Csp*Tm* was previously shown to be described well with a Gaussian chain model over a broad range of denaturant concentrations (17).

[§] The numbers given here indicate the net charge at pH 7.4 of the polypeptide segment between the FRET chromophores (Table S1).

cally at very high denaturant activities. In spite of its simplicity, this model captures the functional form of chain expansion and yields binding constants in agreement with calorimetric measurements (18, 33)[¶]. Charge screening is treated in terms of Debye–Hückel theory, with the Debye length $\kappa^{-1} = (8\pi l_B I)^{-1/2}$, where $I$ is the ionic strength of the solution and $l_B$ is the Bjerrum length, the distance at which the electrostatic energy of the interaction between two elementary charges equals thermal energy, $k_B T$. We obtain $l_B = e^2/(4\pi\varepsilon_0\varepsilon_r k_B T)$, where $e$ is the elementary charge, $\varepsilon_0$ is the permittivity of vacuum, $\varepsilon_r$ is the dielectric constant, $k_B$ is Boltzmann's constant, and $T$ is temperature. To treat the interactions of charges within the chain, we use polyampholyte theory (34), a description that includes both repulsive interactions between charges of the same sign and attractive interactions between charges of opposite sign. Here, we adopt the approach of Higgs and Joanny (35), which can be used to extract Flory-like scaling factors for polyampholytes. This model considers a chain in which monomer $n$ has charge $ec_n$, where $c_n$ assumes the value $+1$ with probability $f$ for the occurrence of a positive charge, $-1$ with the probability $g$ for the occurrence of a negative charge, and 0 with probability $1 - f - g$ for the remaining neutral monomers. $f$ and $g$ at our pH of 7.4 were computed using the p$K_a$ values of all contributing ionizable groups, including those of the fluorophores (Table S1). The monomers have an excluded volume $\nu b^3$ (with a segment length $b = 0.38$ nm, the $C_\alpha$-$C_\alpha$ distance in a polypeptide, and $\nu \geq 0$), and they interact via screened Coulomb interactions between each pair of monomers $m$ and $n$, with distances assumed to follow Gaussian chain statistics. Introducing the expansion factor $\alpha$ as the ratio of an effective segment length $b_1$ and the real segment length $b$, Higgs and Joanny showed that under the condition $N^{1/2}\kappa b_1 > 1$[‖] (where $N$ is the number of segments in the chain), electrostatic interactions can be regarded as a contribution to the effective excluded volume $\nu^* b^3$. The chain dimensions can then be described using the expression

$$\alpha^5 - \alpha^3 = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{3/2} N^{1/2}\nu^*, \quad \text{where } \alpha = \frac{b_1}{b} \qquad [3]$$

and

$$\nu^* b^3 = \nu b^3 + \frac{4\pi l_B (f-g)^2}{\kappa^2} - \frac{\pi l_B^2 (f+g)^2}{\kappa}. \qquad [4]$$

Here, $\nu b^3$ is the excluded volume in the uncharged chain. The second term in Eq. **4** accounts for repulsive interactions due to the net charge of the polypeptide, which result in an increase of $\nu^* b^3$ as in related formulations of polyelectrolyte theory (36) (see *SI Text*). The third term leads to a reduction of $\nu^* b^3$ through attractive interactions between charges of opposite sign. The polyampholyte model can be combined with the binding model [Eq. **2**] to yield

$$R_g = N^{0.5}\alpha b\left(1 + \rho\frac{Ka}{1 + Ka}\right), \qquad [5]$$

which can be used to describe the dependence of $R_g$ on GdmCl activity $a$ (Fig. 4). $\alpha$ is determined by Eqs. **3** and **4**, such that the only fit parameters are $K$, $\rho$, and $\nu$ (see *SI Text*, Table S2, Table S3, Table S4, and Table S5). All other parameters are calculated from the polypeptide sequence and the composition of the solution.

Fits to the GdmCl dependencies of $R_g$ illustrate the strength of this approach (Fig. 4). Polyampholyte theory predicts the rollover in $R_g$ for IN (Fig. 4*B*) and the ProTα variants (Fig. 4 *C* and *D*), both in terms of the functional form and the amplitude of the expansion with good accuracy. Notably, the amplitude does not involve any adjustable fit parameters, but results from the values of $f$ and $g$, which are computed from the charge composition of the polypeptide sequence. Eq. **5** also captures the monotonic collapse of Csp*Tm* (Fig. 4*A*). Surprisingly, it predicts an additional collapse of unfolded Csp*Tm* at low ionic strength caused by the attractive electrostatic interactions between opposite charges in this charge-balanced polyampholyte (34, 35). To illustrate this point, Fig. 4 shows the contributions of the effective binding of GdmCl and of the charge interactions to the changes in $R_g$ as continuous and dashed gray lines, respectively. A charge-induced collapse is consistent with the experimental data, which show a pronounced drop in $R_g$ at the lowest GdmCl concentrations (Fig. 4*A*), but we tested this hypothesis further. If charge attractions indeed contribute to the collapse of unfolded Csp*Tm*, we expect that screening of the charges by adding salt will lead to an expansion of the chain, whereas a collapse is expected for the IDPs. Exactly this behavior is observed experimentally, as Fig. 5 illustrates for ProTαC and Csp*Tm*: if 1.0 M potassium chloride is added to screen the charges, $\langle E\rangle$ of ProTαC increases, corresponding to a collapse (Fig. 4*D*), whereas $\langle E\rangle$ of unfolded Csp*Tm* decreases, indicating an expansion (Fig. 4*A*)[**]. We can thus conclude that charge interactions can indeed lead both to an expansion or a collapse of unfolded proteins, depending on the charge balance in the polypeptide.

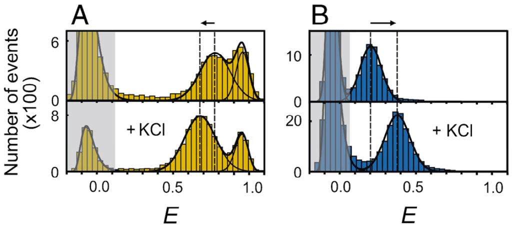

**Fig. 5.** Effect of charge shielding on denatured state dimensions. Transfer efficiency histograms of Csp*Tm* (*A*) and ProTαC (*B*) in 1.0 M urea in the absence (*Upper*) and in the presence of 1.0 M KCl (*Lower*). Csp*Tm* expands on addition of KCl, whereas ProTαC collapses, as predicted by polyampholyte theory. For ProTαC, the change in $\langle E\rangle$ corresponds to a reduction in $R_g$ from (4.2 ± 0.2) nm without KCl to (2.98 ± 0.09) nm with KCl, close to the value of (2.87 ± 0.09) nm calculated with Eq. **5** from the corresponding reduction in Debye length. For unfolded Csp*Tm*, addition of 1.0 M KCl causes an increase in $R_g$ from (1.73 ± 0.05) nm to (1.97 ± 0.05) nm, with a calculated $R_g$ of the charge-shielded unfolded state of (2.15 ± 0.05) nm.

## Discussion

The balance of interactions in the unfolded state and their effect on the compactness of the chain is under intense debate, with possible contributions from hydrophobic interactions (37, 38), hydrogen bonding (39, 40), and charge–charge interactions (21, 41–43). Advances in the application of theoretical models and simulations have addressed important aspects; e.g. the influence of solvation and denaturants (27–29, 44–46), temperature (18), and specific interactions (47–49). With the increasing availability of biophysical methods that can provide structural information even on conformationally heterogeneous systems such as unfolded proteins (1, 12, 43, 50–52), we are starting to be able to correlate experimental findings with theoretical models.

Here we used single-molecule FRET to systematically investigate unfolded proteins with different charge composition, and find that charge interactions play a decisive role for unfolded state dimensions. In proteins with high net charge, as prevalent in many IDPs, charge repulsion can lead to a pronounced expansion (43), similar to previous observations in acid-unfolded proteins (25). Even more strikingly, in proteins with a similar number

---

[¶] Other models can be applied to treat the protein-denaturant interactions, but this has no bearing on our conclusions regarding the electrostatically driven expansion and collapse of the chain we are concerned with in this work.

[‖] The values for $N^{1/2}\kappa b_1$ at the lowest ionic strength of 50 mM used here are 6.8 for ProTαC, 5.5 for ProTαN, 4.1 for IN, and 2.6 for Csp*Tm*.

[**] The experiment is performed at 1.0 M urea to increase the separation of folded and unfolded Csp*Tm* in the transfer efficiency histograms.

of positive and negative charges, charge attraction can amplify the collapse of the chain, a concept that has been well established in polyampholyte theory, but has eluded experimental investigation in the context of proteins (9). We find that polyampholyte theory offers a remarkably good description of the influence of charged amino acids on chain dimensions and thus provides a means of predicting the dimensions of IDPs and unfolded proteins. As expected for polypeptides with a large net charge, the expansion of ProTα and IN at low ionic strength can also be described by polyelectrolyte theory (36, 53–55) (Table S6); i.e. without taking into account attractive charge interactions (Fig. S3). Even though the net charge of Csp*Tm* is close to zero (Table S1), and attractive charge interactions may thus be expected to become dominant, a charge-screened polyelectrolyte in poor solvent could give rise to a collapse behavior similar to what we observe for Csp*Tm* (Fig. 4) (36). However, because the addition of a nonchaotropic salt at low denaturant concentration leads to an expansion of unfolded Csp*Tm* (Figs. 4*A* and 5*A*), attractive charge interactions clearly contribute to its collapse under these conditions. Polyampholyte theory thus appears to be an appropriate generalization of polyelectrolyte theory for describing our results with one simple consistent model. We note, however, that the polyampholyte theory of Higgs and Joanny (35) was developed for polymers in good solvent; i.e., above the theta point. For comparison with the experimentally observed range of values, we estimate the radius of gyration at the theta point from $\langle R_{g\theta}^2\rangle^{3/2} = R_{gN}^3\sqrt{27N/19}$, as suggested by Sanchez (30, 56), where $R_{gN}$ is the radius of gyration of the most compact or folded state. $R_{gN}$ for protein segments of the size investigated here is approximately 1.2 nm, resulting in a radius of gyration of approximately 2.4 nm at the theta point. Only unfolded Csp*Tm* in low concentrations of denaturant compacts much below this value of $R_g$, but it still remains more expanded than the native state. The expansion of collapsed unfolded Csp*Tm* on addition of KCl indicates the importance of attractive charge interactions even under these conditions, but it will be interesting to assess the limitations of our analysis, such as the additional role of three-body interactions, effects from the finite size of the polypeptides, and possible correlations between charges within the chain, with more advanced theoretical approaches and molecular simulations (9, 43, 57).

The range of ionic strengths where we observe the expansion of IDPs corresponds to physiologically relevant values, indicating the importance of the effect in a cellular environment. The resulting changes in chain dimensions will affect many of the properties that characterize IDP function, including the capture radii for fly-casting (5, 6), repulsive entropic forces from brush-like structures (9, 10), and the free energy change of folding upon binding (2). From our results, we can estimate the repulsive electrostatic energy that needs to be overcome to compact the chain compared to a neutral polypeptide. At a physiologically relevant ionic strength of 100 mM, polyampholyte theory yields repulsive electrostatic energies[††] of $(1.4 \pm 0.5)\ k_BT$ and $(4.2 \pm 0.5)\ k_BT$ for the two variants of ProTα, a negligible value of $(-0.2 \pm 0.5)\ k_BT$ for IN, and an attractive energy of $(-2.5 \pm 0.5)\ k_BT$ for Csp*Tm*, underlining the critical balance of positive and negative charges that leads to a transition from a pronounced expansion to a charge-mediated compaction of the polypeptides (Fig. S4). Considering the magnitude of the values, the additional energy required to compact and fold IDPs with a large net charge can be a major component that has to be overcompensated by the free energy of binding to their target molecules. The electrostatic energies are in a range sufficient for significantly modulating affinities and dynamics of intracellular interaction partners, thus allowing for high specificity, while at the same time enabling dissociation at a rate that ensures a response of the regulatory system on time scales relevant for cellular processes (2).

Our results also have important implications for the properties of unfolded proteins in general. A large number of experiments have demonstrated a collapse of unfolded proteins at low denaturant concentration (12, 59, 60). However, the interactions responsible for this collapse have been difficult to elucidate. Correspondingly, it has been unclear how collapse is affected by sequence composition, and especially whether IDPs exhibit a similar collapse behavior as the previously investigated globular proteins. Some experiments, theoretical considerations, and simulations have suggested a role of secondary structure formation and hydrogen bonding for collapse (18, 39, 40), even though this is improbable to be the only contribution (45). Recent findings that IDPs and unstructured peptides can form collapsed structures (39, 61, 62) and show a compaction with increasing temperature similar to globular unfolded proteins (18) may be surprising given the lack of hydrophobic side chains and have been taken to indicate that hydrophobic interactions are not predominant in determining unfolded state dimensions. Our results show that charge interactions can play a decisive role. The observation of a charge-driven expansion even for a nearly charge-balanced protein like IN indicates that this behavior will be relevant for a large number of proteins, as recently suggested from results on barstar (42). Investigations of charged IDPs and unfolded states of globular proteins will thus depend critically on the solution conditions, such as pH and salt concentration (21, 25, 63–65).

In summary, whereas our results indicate a denaturant-dependent collapse of IDPs qualitatively similar to that of unfolded globular proteins at high ionic strength, charge interactions can dominate the dimensions of IDPs at low ionic strength. Our findings provide a new opportunity for testing the suitability of polymer physical concepts for describing unfolded state behavior. Interestingly, already a simple polyampholyte theory captures the overall effect of charge interactions on unfolded proteins remarkably well. Together with previous results on the stiffness of uncharged chains (12, 66), this presents a new possibility to predict the effect of the content of charged amino acid on the dimensions of unfolded proteins and IDPs, and to explore the resulting impact on protein stability and interactions. Whereas unfolded proteins have been shown to follow simple scaling laws at high concentrations of GdmCl (26), similar to the behavior expected for homopolymers, sequence composition can obviously have a large effect on chain dimensions under physiological conditions, with important implications for protein dynamics and interactions.

## Methods

**Preparation and Labeling of Proteins.** Cys residues were introduced by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes Alexa Fluor 488 and 594 essentially as described previously (17, 18, 23). All proteins were purified using a hexahistidine tag. For details, see ref. 18 and *SI Text*.

**Single-Molecule Fluorescence Spectroscopy.** Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) essentially as described previously (17). Samples were measured with a 20 to 50 pM protein concentration in 50 mM Tris buffer, pH 7.4. 0.001% Tween 20 (Pierce) was included to prevent surface adhesion of the proteins. IN devoid of bound $Zn^{2+}$ was prepared by adding 1 mM EDTA. To minimize photochemical damage to the chromophores, 200 mM β-mercaptoethanol were included in the samples (18). The Förster radius $R_0$ was corrected for the changes in solution conditions, which were dominated by the change in refractive index with GdmCl, urea, and KCl concentration. All errors given are our estimates of the experimental uncertainty; wherever possible, they represent standard deviations from multiple independent meaurements. For details, see *SI Text*.

**ACKNOWLEDGMENTS.** We thank A. Vartapetian and A. Evstafieva for a plasmid encoding prothymosin α, R. Craigie for a plasmid encoding the N-terminal domain of HIV integrase and helpful discussion. We thank Robert Best, Dimitrii

---

[††] Electrostatic energies were calculated according to $\frac{\Delta F}{k_BT} = \frac{1}{2}\frac{N^{0.5}\nu b^3}{b_1^3}$ (58).

Makarov, Devarajan Thirumalai, Vladimir Uversky, and Andreas Vitalis for very helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation, the National Center of Competence in Research for Structural Biology, a Starting Researcher Grant of the European Research Council, and a fellowship of the Deutscher Akademische Austauschdienst (V.H.).

1. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
2. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
3. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18:756–764.
4. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: Introducing the D2 concept. *Ann Rev Biophys* 37:215–246.
5. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci USA* 97:8868–8873.
6. Turjanski AG, Gutkind JS, Best RB, Hummer G (2008) Binding-induced folding of a natively unstructured transcription factor. *PLoS Comput Biol* 4:e1000060.
7. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: Conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 93:11504–11509.
8. Oldfield CJ, et al. (2008) Flexible nets: Disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* 9(Suppl 1):S1.
9. Bright JN, Woolf TB, Hoh JH (2001) Predicting properties of intrinsically unstructured proteins. *Prog Biophys Mol Biol* 76:131–173.
10. Brown HG, Hoh JH (1997) Entropic exclusion by neurofilament sidearms: A mechanism for maintaining interfilament spacing. *Biochemistry* 36:15035–15040.
11. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
12. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
13. Haran G (2003) Single-molecule fluorescence spectroscopy of biomolecular folding. *J Phys-Condens Mat* 15:R1291–R1317.
14. Michalet X, Weiss S, Jäger M (2006) Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem Rev* 106:1785–1813.
15. Deniz AA, et al. (2000) Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA* 97:5179–5184.
16. Nettels D, Gopich IV, Hoffmann A, Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci USA* 104:2655–2660.
17. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
18. Nettels D, et al. (2009) Single molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci USA* 106:20740–20745.
19. Cai M, et al. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol* 4:567–577.
20. Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* 93:13659–13664.
21. Gast K, et al. (1995) Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* 34:13211–13218.
22. Pineiro A, Cordero OJ, Nogueira M (2000) Fifteen years of prothymosin alpha: Contradictory past and new horizons. *Peptides* 21:1433–1446.
23. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
24. Magg C, Schmid FX (2004) Rapid collapse precedes the fast two-state folding of the cold shock protein. *J Mol Biol* 335:1309–1323.
25. Uversky VN, Goto Y (2009) Acid denaturation and anion-induced folding of globular proteins: Multitude of equilibium partially folded intermediates. *Curr Protein Pept Sc* 10:447–455.
26. Kohn JE, et al. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
27. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
28. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
29. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11:83–93.
30. Ziv G, Haran G (2009) Protein folding, protein collapse, and Tanford's transfer model: Lessons from single-molecule FRET. *J Am Chem Soc* 131:2942–2947.
31. Flory PJ, Fisk S (1966) Effect of volume exclusion on dimensions of polymer chains. *J Chem Phys* 44:2243–2248.
32. Bright JN, Woolf TB, Hoh JH (2001) Predicting properties of intrinsically unstructured proteins. *Prog Biophys Mol Biol* 76:131–173.
33. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226:491–505.
34. Dobrynin AV, Colby RH, Rubinstein M (2004) Polyampholytes. *J Polym Sci Pol Phys* 42:3513–3538.
35. Higgs PG, Joanny JF (1991) Theory of Polyampholyte Solutions. *J Chem Phys* 94:1543–1554.
36. Ha BY, Thirumalai D (1992) Conformations of a polyelectrolyte chain. *Phys Rev A* 46:R3012–R3015.
37. Hiller S, Wider G, Imbach LL, Wuthrich K (2008) Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX. *Angew Chem Int Ed Engl* 47:977–981.
38. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA* 105:6278–6283.
39. Möglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103:12394–12399.
40. Bolen DW, Rose GD (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* 77:339–362.
41. Zhou HX (2002) A Gaussian-chain model for treating residual charge–charge interactions in the unfolded state of proteins. *Proc Natl Acad Sci USA* 99:3569–3574.
42. Hofmann H, Golbik RP, Ott M, Hübner CG, Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* 376:597–605.
43. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:8183–8188.
44. O'Brien EP, Ziv G, Haran G, Brooks BR, Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc Natl Acad Sci USA* 105:13403–13408.
45. Tran HT, Mao A, Pappu RV (2008) Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J Am Chem Soc* 130:7380–7392.
46. Vitalis A, Wang X, Pappu RV (2007) Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys J* 93:1923–1937.
47. Kohn JE, Gillespie B, Plaxco KW (2009) Non-sequence-specific interactions can account for the compaction of proteins unfolded under "native" conditions. *J Mol Biol* 394:343–350.
48. Weinkam P, Pletneva EV, Gray HB, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles. *Proc Natl Acad Sci USA* 106:1796–1801.
49. Shoemaker BA, Wolynes PG (1999) Exploring structures in protein folding funnels with free energy functionals: The denatured ensemble. *J Mol Biol* 287:657–674.
50. Nodet G, et al. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc* 131:17908–17918.
51. Millett IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv Protein Chem* 62:241–262.
52. Gast K, Modler AJ (2005) Studying protein folding and aggregation by laser light scattering. *Protein Folding Handbook*, eds J Buchner and T Kiefhaber (Wiley-VCH, Weinheim), pp 673–709.
53. Odijk T (1977) Polyelectrolytes near the Rod Limit. *J Polym Sci Pol Phys* 15:477–483.
54. Skolnick J, Fixman M (1977) Electrostatic persistence length of a wormlike polyelectrolyte. *Macromolecules* 10:944–948.
55. Ha BY, Thirumalai D (1999) Persistence length of flexible polyelectrolyte chains. *J Chem Phys* 110:7533–7541.
56. Sanchez IC (1979) Phase-transition behavior of the isolated polymer-chain. *Macromolecules* 12:980–988.
57. Dobrynin AV, Rubinstein M (1995) Flory theory of a polyampholyte chain. *J Phys II France* 5:677–695.
58. Chan HS, Dill KA (1991) Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 20:447–490.
59. Uversky VN (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* 32:13288–13298.
60. Uversky VN, Ptitsyn OB (1994) "Partly folded" state, a new equilibrium state of protein molecules: Four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature. *Biochemistry* 33:2782–2791.
61. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV (2006) Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci USA* 103:16764–16769.
62. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA (2007) A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. *Proc Natl Acad Sci USA* 104:2649–2654.
63. Uversky VN, et al. (1999) Natively unfolded human prothymosin alpha adopts partially folded collapsed conformation at acidic pH. *Biochemistry* 38:15009–15016.
64. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR (2007) Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 367:609–615.
65. Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28:305–325.
66. Zhou HX (2004) Polymer models of protein stability, folding, and interactions. *Biochemistry* 43:2141–2154.
67. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132.
68. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single molecule fluorescence. *Proc Natl Acad Sci USA* 102:2754–2759.

# Supporting Information

## Müller-Späth et al. 10.1073/pnas.1001743107

### SI Text

**SI Methods.** ***Preparation and labeling of proteins.*** Cysteine residues were introduced by site-directed mutagenesis to provide functional groups for the specific attachment of the dyes as described previously (1, 2). The truncated variant of Csp*Tm* was expressed with a cleavable hexahistidine tag to allow for rapid purification. The gene was cloned from vector pET21a (1, 2) into pET47b(+) and the sequence coding for the two C-terminal amino acids was deleted by site-directed mutagenesis. The protein was expressed in LB medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were disrupted and DNA was digested. The supernatant was cleared by centrifugation and loaded on a HisTrap column (GE Healthcare, BioSciences AB) in 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, 4 M GdmCl, pH 8.0. After the 280 nm UV absorption signal reached the baseline, the column was washed with two column volumes 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 8.0, and a gradient from 10 to 500 mM imidazole was used to elute the His-tagged protein. HRV 3C protease (containing a His-tag) was added to a final concentration of 0.3 mg/mL, and after 12 h at room temperature, the cleavage reaction was dialyzed against 20 mM Tris-HCl, 0.5 M NaCl, 2 mM β-mercaptoethanol, 10 mM imidazole, pH 8.0 and applied to a HisTrap column. The cleaved Csp*Tm* without His-tag was collected in the flow-through and concentrated. Labeling was performed as described previously (2).

The N-terminal domain of HIV1-integrase (IN) was expressed in the vector pET15b. Cysteine residues were introduced at positions 0 and 56 (residue numbering is starting from Phe in the protein sequence; Table S1). The protein was expressed in LB medium with carbenicillin and 1 mM IPTG at 37 °C. Harvested cells were lysed and the DNA was digested. The supernatant was cleared by centrifugation and loaded on a HisTrap HP column, equilibrated with 20 mM HEPES, 1 M sodium chloride, 20 mM imidazole, 2 mM β-mercaptoethanol, pH 7.5. The column was washed with 60 mM imidazole and IN was eluted using 10% glycerol and a gradient from 20 to 500 mM imidazole maintaining all other buffer conditions. Fractions were identified via SDS-PAGE, combined and dialyzed against 20 mM HEPES, 100 mM sodium chloride, 5 mM EDTA, 1 mM DTT, pH 7.5. The His-tag was cleaved using 10 units thrombin from bovine plasma (SERVA Electrophoresis, Heidelberg, Germany) per mg protein at an IN concentration of 0.13 mg/mL at room temperature for 30 minutes. Cleavage was verified by SDS-PAGE, and the protein was loaded on the HisTrap HP column under the same conditions as before. Fractions containing IN were combined, adjusted to 6 M GdmCl and 3 mM TCEP, and concentrated by ultrafiltration. Gel filtration was done with a HiLoad Superdex 75 prep grade column (GE Healthcare, BioSciences AB) under refolding conditions in 25 mM Tris, 250 mM sodium chloride, 10% glycerol, 0.1 mM $ZnCl_2$, 1 mM DTT, pH 7.4. For fluorophore labeling, IN was reduced with 3 mM TCEP and desalted with a HiTrap desalting column (GE Healthcare, BioSciences AB) in 50 mM sodium phosphate, 0.1 mM $ZnCl_2$ pH 7.0. Fractions were collected under argon atmosphere. The protein was incubated with Alexa Fluor 488 maleimide at a 1 : 1 molar ratio. Singly labeled protein was separated from unlabeled and doubly labeled protein using ion exchange chromatography with a MonoQ column (GE Healthcare, BioSciences AB). Fractions containing singly labeled protein, confirmed by matrix-assisted laser desorption/ionization mass spectrometry, were incubated with a 2 : 1 molar excess of Alexa Fluor 594 maleimide. Doubly labeled protein was separated as before and the correct molecular mass of the labeled protein was confirmed by mass spectrometry.

The coding sequence for human ProTα was cloned from vector pHP12 (3) into pET47b(+). Cysteine residues were introduced in positions 2 and 56 in the variant ProTαN and in positions 56 and 110 in the variant ProTαC by site-directed mutagenesis (residue numbering is from Met in the protein sequence, excluding the 19 residue N-terminal purification tag; Table S1). The protein was expressed in Terrific Broth medium with kanamycin and 1 mM IPTG at 37 °C. Harvested cells were lysed and DNA was digested. The cleared supernatant was loaded on a HisTrap HP column (GE Healthcare, BioSciences AB) in 20 mM Tris, 100 mM sodium chloride and 2 mM β-mercaptoethanol, pH 7.0. A gradient from 40 to 500 mM imidazole was used to elute the His-tagged protein. Fractions were identified via SDS-PAGE, combined, extracted by butanol, and precipitated with ethanol (3). Pellets were dissolved in 50 mM sodium phosphate, pH 7.0, reduced with 5 mM TCEP, and purified on a Superdex75 gel filtration column (GE Healthcare, BioSciences AB) in 100 mM sodium phosphate, 2 mM β-mercaptoethanol, 0.01% Tween, pH 7.0. Fractions containing the full-length protein were combined, extracted with butanol, and precipitated with ethanol. The pellet was dissolved in 4 M GdmCl, 50 mM sodium phosphate, pH 7.0, and the protein concentration was determined with a bicinchoninic acid assay (BCA Protein Assay Kit, Pierce), because ProTα contains no aromatic residues. Fluorophore labeling was performed at a protein concentration of approximately 0.1 mg/mL with a threefold molar excess of the dyes Alexa Fluor 488 and Alexa Fluor 594 maleimide; gel filtration was used to remove the free dye. The correct molecular mass of the labeled protein was confirmed by matrix-assisted laser desorption/ionization mass spectroscopy. The donor-only labeled and acceptor-only labeled ProTα resulting from such random labeling does not interfere with single-molecule FRET measurements.

***Single-molecule fluorescence spectroscopy.*** Observations of single-molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) equipped with a continuous wave 488 nm diode laser (Sapphire 488-100 CDRH, Coherent) and an Olympus UplanApo 60x/1.20W objective. Sample fluorescence was separated into donor and acceptor components using a dichroic mirror (585DCXR, Chroma), and two final filters (Chroma ET525/50M, HQ650/100). Each component was focused onto an avalanche photodiode (SPCM-AQR-15, PerkinElmer Optoelectronics), and the arrival time of every detected photon was recorded. Samples of labeled protein were diluted to a concentration of approximately 20 pM in 50 mM Tris buffer at the appropriate GdmCl (Pierce) concentration, and individually adjusted to pH 7.4. 0.001% Tween 20 (Pierce) was added to prevent surface adhesion of the protein (1). To minimize damage to the chromophores, the photo-protective additive β-mercaptoethanol (200 mM) was included. To eliminate zinc from its complex with IN, 1 mM EDTA was added. The measurements were performed at a laser power of 110 μW at the sample with an acquisition time of 1 h (for 7000 to 15000 identified bursts). Successive photons detected in either channel separated by less than 100 μs were combined into one burst. Identified bursts were corrected for background, differences in quantum yields of donor and acceptor, the different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation as described previously (4). A burst was retained as a significant event if the total number of counts exceeded 50.

The Förster radius $R_0$ was corrected for the changes in solution conditions, which were dominated by the change in refractive index with GdmCl (2, 5) or urea (6, 7) concentration. The overlap integral (8) of Alexa 488 emission and Alexa 594 absorption was found to be independent of denaturant concentration.

***Determination of the radius of gyration.*** We determined the dimensions of the unfolded protein chains from the measured transfer efficiency with two different approaches: the Gaussian chain (1, 2, 9–11) and a variation of Sanchez theory previously used by Haran and coworkers (10, 12, 13).

The end-to-end distance distribution of a Gaussian chain is

$$P_{\text{Gaussian}}(r) = 4\pi r^2 \left(\frac{3}{2\pi\langle r^2\rangle}\right)^{3/2} \exp\left(-\frac{3r^2}{2\langle r^2\rangle}\right), \qquad \textbf{[S1]}$$

and its mean-square radius of gyration is given by $\langle R_g^2\rangle = \langle r^2\rangle/6$.

The Sanchez-type theory (10, 12, 13) employs the modified Flory–Fisk equation (12, 13)

$$P(R_g) = P_0(R_g)\exp(-Ng(\phi,\varepsilon)/\mathrm{k_B T}), \qquad \textbf{[S2]}$$

where $P_0(R_g) \propto R_g^6 \exp(-7/2\; R_g^2/\langle R_{g\theta}^2\rangle)$ is the Flory–Fisk distribution for the radius of gyration of an ideal polymer chain (with a normalization chosen such that the integral of $P(R_g)$ equals unity), which is weighted by the expansion free energy per monomer, $g(\phi,\varepsilon) = -\frac{1}{2}\phi\varepsilon + \mathrm{k_B T}\frac{1-\phi}{\phi}\log(1-\phi)$. The function $g(\phi,\varepsilon)$ is defined according to eq. 26 of ref. 14, neglecting the constant term that can be included in the normalization factor. $\phi = R_{g,N}^3/R_g^3$ is the volume fraction occupied by the chain ($R_{g,N}$ is the radius of gyration of the fully compact/native state), and $\varepsilon$ is a mean field interaction relative to the most collapsed state and is a measure of the two-body interactions within the chain. The radius of gyration at the $\theta$-point can be estimated using the argument of Sanchez (13, 14) based on Landau's theory of phase transitions (13). Accordingly, the volume fraction in the $\theta$-state depends on the number of amino acids, $N$, as $\phi_\theta = \sqrt{19/27}N^{-1/2}$. With the radius of gyration of a (hypothetical) native state corresponding to protein segments of the size investigated here (calculated by scaling the radius of gyration of Csp*Tm*, $R_{g,N0} \approx 1.2$ nm, for the number of amino acids $N$ in the segment as $R_{g,N}^3 = R_{g,N_0}^3 N/N_0$), the radius of gyration of the $\theta$-state is approximately 2.4 nm.

We convert the distribution for the radius of gyration into a distribution of the end-to-end distance using the approximation (13)

$$P_{\text{Sanchez}}(r) = \int p(r|R_g)P_{\text{Sanchez}}(R_g)dR_g, \qquad \textbf{[S3]}$$

where $p(r|R_g)$ is the conditional probability for a distribution of end-to-end distances given a value of $R_g$. $p(r|R_g)$ is taken as the distance distribution of two random points inside a sphere with a corrected radius of gyration $\delta \cdot R_g$:

$$p(r|R_g) = \frac{1}{\delta\cdot R_g}\left(3\left(\frac{r}{\delta\cdot R_g}\right)^2 - \frac{9}{4}\left(\frac{r}{\delta\cdot R_g}\right)^3 + \frac{3}{16}\left(\frac{r}{\delta\cdot R_g}\right)^5\right), \qquad \textbf{[S4]}$$

where the phenomenological factor δ is chosen such that $\langle R_g^2\rangle = \langle r^2\rangle/6$ at the $\theta$-point with Eq. **S3** and $P_{\text{Sanchez}}(r) = P_0(R_g)$, resulting in $\delta = 2.26$ in our case. Finally, $P_{\text{Sanchez}}(r)$ is used to fit the experimental mean FRET efficiencies by adjusting $\varepsilon$ (Fig. S2). The average radius of gyration is then calculated from the resulting $P(R_g)$.

A comparison between the values of radius of gyration obtained with the two different treatments is reported in Fig. S1: significant differences are only observed in the case of the two variants of ProTα at very low ionic strength, where the chain is expanded far beyond the estimated $\theta$-point, but this does not affect our conclusions significantly.

***Polyelectrolyte theory.*** The simplest description that takes into account electrostatic repulsion considers the unfolded protein a polyelectrolyte, i.e. a polymer with only one type of charge (15–18). Ha and Thirumalai (18) showed that the effect of charges on the conformations of a polyelectrolyte chain can be described in terms of an effective excluded volume. Calculating the free energy and the dimensions of the chain by a standard self-consistent variational treatment, Ha and Thirumalai described the expansion factor α as

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} - f_{\text{el}}(\kappa)\alpha^2 = X, \qquad \textbf{[S5]}$$

where $y$ provides an estimate of the three-body interaction, $f_{\text{el}}(\kappa)$ describes the electrostatic interactions as a function of the Debye screening length $\kappa^{-1}$, and $X$ is related to the monomer excluded volume $vb^3$ through

$$X = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} vN^{0.5}. \qquad \textbf{[S6]}$$

In the limit of $\kappa b\alpha \gg 0$ in the presence of salt, the equation can be rewritten as

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} vN^{0.5} + 2\left(\frac{6}{\pi}\right)^{0.5} uZ^2N^{0.5}\frac{1}{(\kappa b)^2}, \qquad \textbf{[S7]}$$

where $u = l_B/b$ is the ratio of the Bjerrum length and the monomer length, and $Z$ is the charge per monomer. Interpreting $Z$ as the density of net charge over the chain and considering $f$ and $g$ as the fractions of positive and negative charges per monomer, $Z = f - g$, we obtain

$$\alpha^5 - \alpha^3 - \frac{y}{\alpha^3} = \frac{4}{3}\left(\frac{3}{2\pi}\right)^{1.5} N^{0.5}\left(v + 2\pi\frac{l_B(f-g)^2}{\kappa^2 b^3}\right). \qquad \textbf{[S8]}$$

As pointed out by Ha and Thirumalai, the right side of the equation can be considered an effective excluded volume term. This result is formally equivalent to neglecting attractive interactions in Eq. 4, such that the excluded volume can be expressed as

$$v^*b^3 = vb^3 + \frac{4\pi l_B(f-g)^2}{\kappa^2}. \qquad \textbf{[S9]}$$

The difference by a factor of 2 in the electrostatic term derives from slightly different approximations adopted in the calculations. The factor of 4 reported by Higgs and Joanny is also found by Muthukumar (19). We adopt this value for simplifying the comparison between polyelectrolyte and polyampholyte theory.

With this formalism, the dimensions of a polyelectrolyte chain can be analyzed analogous to polyampholyte theory in combination with Eqs. 3 and 5. The fits show that Eq. **S1** is a good approximation for ProTα, with its large proportion of glutamate and aspartate residues, and captures the rollover of $R_g$ at low ionic strength. It still correctly predicts a small expansion of the chain for IN at low GdmCl concentrations. To fit the data of Csp*Tm*, however, it is necessary to compensate the effect of the electrostatic term by adjusting the solvent quality. In Fig. S3, we report the predicted rollover when a change in solvent quality is neglected, i.e. if we assume a constant value for the parameter $v$. As pointed out in the *Discussion* of the main text, the KCl addition experiments (Fig. 5) indicate that in the case of Csp*Tm* (where there is not a clear dominance of one type of charge), attractive interactions between opposite charges within the polypeptide contribute to chain collapse. The sequences and charge distribution of all proteins investigated here are given in Table S1.

***Results and Discussion.*** The values of all parameters obtained from fits with Eq. 5 to the results obtained with the Gaussian chain model and the modified Sanchez theory (Fig. S1) are given in Table S2, Table S3, Table S4, and Table S5. The effective binding constants for GdmCl of the protein variants investigated (Table S2, Table S3, Table S4, and Table S5) are between 0.2 and 1.3, in the range of binding constants reported previously (20–22), with the value for Csp*Tm* being slightly higher than for IN and ProTα. Similar results are obtained from fits of Eq. 2 to the urea data (Table S4 and Table S5). Considering that the preferential interaction of denaturants with the polypeptide is expected to depend on the nature of the side chains (21–28), e.g. through the complexation of guanidinium ions by acidic side chains (29), some variation in the effective binding constant with sequence composition may not be surprising. The values for the excluded volume $\nu b^3$ are similar for Csp*Tm* and IN; the larger value for ProTα is probably at least partially due to the presence of the large unlabeled segments in the polypeptide, which will exert an excluded volume effect on the labeled part.

1. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419:743–747.
2. Hoffmann A, et al. (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104:105–110.
3. Evstafieva AG, et al. (1995) Overproduction in *Escherichia coli*, purification and properties of human prothymosin alpha. *Eur J Biochem* 231:639–643.
4. Schuler B (2007) Application of single-molecule Förster resonance energy transfer to protein folding. *Methods Mol Biol* 350:115–138.
5. Nozaki Y (1972) The preparation of guanidine hydrochloride. *Methods Enzymol* 26 PtC:43–50.
6. Warren JR, Gordon JA (1966) On the refractive indices of aqueous solutions of urea. *J Phys Chem* 70:297–300.
7. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* 131:266–280.
8. Van Der Meer BW, Coker G III, Chen S-YS (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).
9. O'Brien EP, Morrison G, Brooks BR, Thirumalai D (2009) How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? *J Chem Phys* 130:124903.
10. Sherman E, Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103:11539–11543.
11. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
12. Ziv G, Thirumalai D, Haran G (2009) Collapse transition in proteins. *Phys Chem Chem Phys* 11:83–93.
13. Ziv G, Haran G (2009) Protein folding, protein collapse, and Tanford's transfer model: Lessons from single-molecule FRET. *J Am Chem Soc* 131:2942–2947.
14. Sanchez IC (1979) Phase-transition behavior of the isolated polymer-chain. *Macromolecules* 12:980–988.
15. Odijk T (1977) Polyelectrolytes near the rod limit. *J Polym Sci Pol Phys* 15:477–483.
16. Skolnick J, Fixman M (1977) Electrostatic persistence length of a worm-like polyelectrolyte. *Macromolecules* 10:944–948.
17. Ha BY, Thirumalai D (1999) Persistence length of flexible polyelectrolyte chains. *J Chem Phys* 110:7533–7541.
18. Ha BY, Thirumalai D (1992) Conformations of a polyelectrolyte chain. *Phys Rev A* 46:R3012–R3015.
19. Muthukumar M (1987) Adsorption of a polyelectrolyte chain to a charged surface. *J Chem Phys* 86:7230–7235.
20. Makhatadze GI, Privalov PL (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study. *J Mol Biol* 226:491–505.
21. Tanford C (1970) Protein denaturation. Part C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* 24:1–95.
22. Schellman JA (2002) Fifty years of solvent denaturation. *Biophys Chem* 96:91–101.
23. Wallqvist A, Covell DG, Thirumalai D (1998) Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *J Am Chem Soc* 120:427–428.
24. Caflisch A, Karplus M (1999) Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure* 7:477–488.
25. Bennion BJ, Daggett V (2003) The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci USA* 100:5142–5147.
26. Stumpe MC, Grubmuller H (2007) Interaction of urea with amino acids: Implications for urea-induced protein denaturation. *J Am Chem Soc* 129:16126–16131.
27. Auton M, Holthauzen LM, Bolen DW (2007) Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc Natl Acad Sci USA* 104:15317–15322.
28. Nodet G, et al. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J Am Chem Soc* 131:17908–17918.
29. O'Brien EP, Dima RI, Brooks B, Thirumalai D (2007) Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: Lessons for protein denaturation mechanism. *J Am Chem Soc* 129:7346–7353.

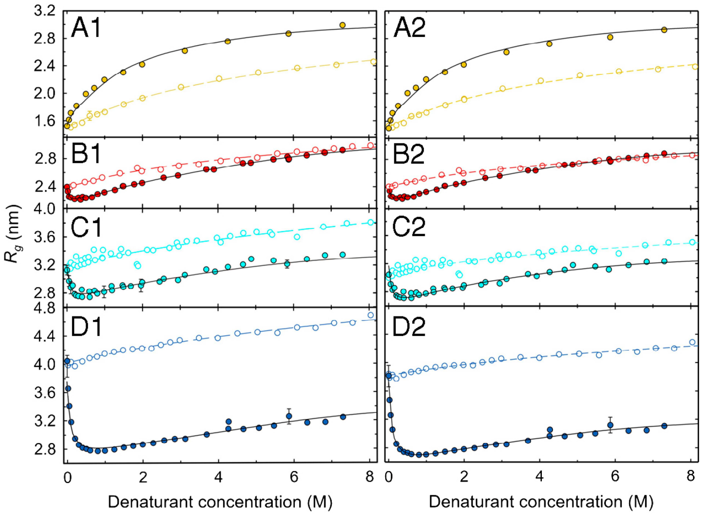

**Fig. S1.** Comparison of the apparent root-mean-square radii of gyration of the labeled protein segments using a Gaussian chain model (Eq. **S1**, *Left*, same fits as in Fig. 4) or the modified Sanchez theory (Eq. **S3**, *Right*). Data are reported as a function of GdmCl (filled circles) and urea (open circles) concentration, with (*A*) Csp*Tm* (yellow), (*B*) IN (red), (*C*) ProTαN (cyan), and (*D*) ProTαC (blue). Fits according to polyampholyte theory (Eq. **5**) are shown as lines. Note that the fits are performed based on thermodynamic activities, but plotted on a concentration scale.

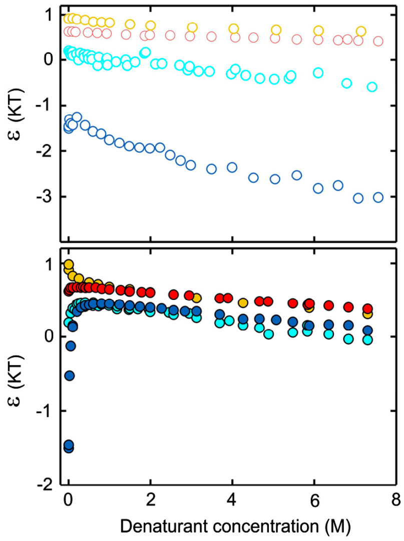

**Fig. S2.** Dependence of the effective interaction parameter $\varepsilon$ in the modified Sanchez theory [Eq. **S2**] on the concentration of GdmCl (*Lower*) and urea (*Upper*) for Csp*Tm* (yellow), IN (red), ProTαN (cyan), and ProTαC (blue).

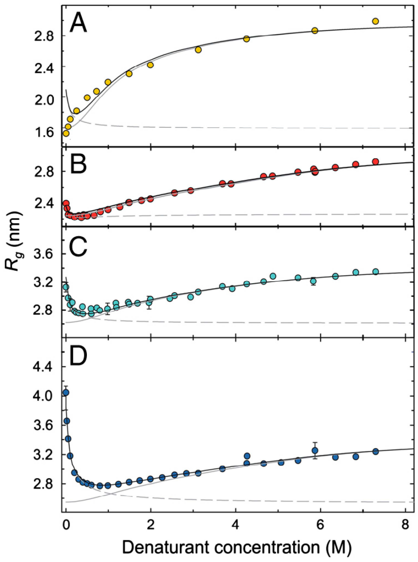

**Fig. S3.** Dependence of the apparent radii of gyration ($R_g$) of the labeled protein segments on the concentration of GdmCl, with (*A*) Csp*Tm* (yellow), (*B*) IN (red), (*C*) ProTαN (cyan), and (*D*) ProTαC (blue). In contrast to Fig. 4, fits according to polyelectrolyte theory with a fixed excluded volume (Eq. **S2**, black solid line) are shown. The two components of the fit to Eq. **S2**, corresponding to GdmCl binding and electrostatic repulsion, are indicated as continuous and dashed gray lines, respectively. The rollover in the fit of Csp*Tm* could be eliminated by a compensating pronounced variation in the excluded volume term $\nu$ [Eq. **S9**], corresponding to a strong change in solvent conditions over this range of GdmCl concentration. Note that the fits are performed based on thermodynamic activities, but plotted on a concentration scale.

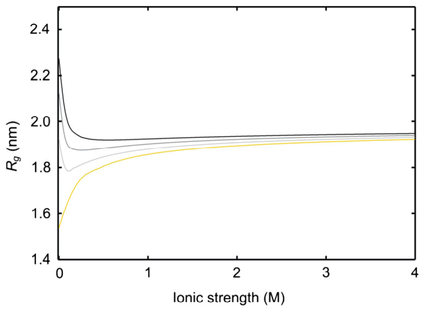

**Fig. S4.** Transition from collapse to expansion at low ionic strength with changing charge balance, illustrating the pronounced sensitivity for charge composition of the chain. The dependencies are calculated according to Eq. 5 with the following parameters: excluded volume $\nu b^3 = 0.2$ nm$^3$, number of segments $N = 53$, number of negatively charged segments $N_- = 14$, number of positively charged segments $N_+ = 8$ (Csp*Tm*, yellow line), $N_+ = 7$ (light gray line), $N_+ = 6$ (dark gray line), $N_+ = 5$ (black line).

**Table S1. Sequences, charges, and labeling positions of the proteins investigated**

| A | B | C | D |
|---|---|---|---|
| Csp*Tm* | −2 | −6 | 1 10 20 30 40 50 54 60 64<br>GPG **C**RGKVKWFDS KKGYGFITKD EGGDVFVHWS AIEMEGFKTL KEGQVVEFEI QEG**C**KGGQAA HVKV<br>+ + + − ++ +− − − − − + +− − − − + + |
| IN | −4 | −8 | 0 1 10 20 30 40 50 56<br>GSH**C** FLDGIDKAQE EHEKYHSNWR AMASDFNLPP VVAKEIVASC DKCQLKGEAM HGQVD**C**<br>− −+ − − −+ + − +− −+ + − − |
| ProTαN<br>(C2–C56) | −14 | −18 | 2 10 20 30 40 50<br>MAHHHHHHS AALEVLFQGP M**S**DAAVDTSS EITTKDLKEK KEVVEEAENG RDAPANGNAN EENGEQEADN<br>− − − − +− +−+ +− −− − +− −− − − − |
| ProTαC<br>(C56–C110) | −27 | −31 | 56 60 70 80 90 100 110<br>EVDEE**C**EEGG EEEEEEEEGD GEEEDGDEDE EAESATGKRA AEDDEDDDVD TKKQKTDED**D**<br>− −−− −− −−−−−−−− − −−−− −−−− − − ++ −−−−−−− − ++ + −−−− |

(A) Protein, (B) Net charge of interdye sequence, (C) Net charge of interdye sequence including the charges of the dyes, and (D) Amino acid sequence and positions where Cys residues were introduced for labeling (indicated in **bold**); charges at pH 7.4 are indicated below the respective amino acids.

**Table S2. GdmCl concentration dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with polyampholyte theory (Eq. 5)**

|  | $K$ | $\nu b^3$ (nm$^3$) | $\rho$ |
|---|---|---|---|
| Csp*Tm* | 1.3 ± 0.2* | 0.20 ± 0.01 | 0.56 ± 0.02 |
| IN | 0.34 ± 0.03 | 0.52 ± 0.01 | 0.48 ± 0.02 |
| ProTαN | 0.4 ± 0.2 | 1.14 ± 0.06 | 0.28 ± 0.04 |
| ProTαC | 0.2 ± 0.1 | 1.26 ± 0.08 | 0.29 ± 0.09 |

*The errors given represent only the uncertainty of the fit

**Table S3. GdmCl concentration dependence obtained with Eq. S3 (Sanchez model, Fig. S1); parameters from fits with polyampholyte theory (Eq. 5)**

|  | $K$ | $\nu b^3$ (nm$^3$) | $\rho$ |
|---|---|---|---|
| Csp*Tm* | 1.3 ± 0.2 | 0.20 ± 0.01 | 0.56 ± 0.02 |
| IN | 0.43 ± 0.03 | 0.38 ± 0.01 | 0.32 ± 0.01 |
| ProTαN | 0.4 ± 0.1 | 1.02 ± 0.04 | 0.24 ± 0.02 |
| ProTαC | 0.5 ± 0.1 | 0.91 ± 0.05 | 0.25 ± 0.02 |

**Table S4. Urea concentration dependence dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with binding model (Eq. 2)**

|  | $K$ | $R_{g0}$ (nm) | $\rho$ |
|---|---|---|---|
| Csp*Tm* | 0.20 ± 0.02 | 1.47 ± 0.01 | 1.15 ± 0.06 |
| IN | 0.13 ± 0.02 | 2.39 ± 0.01 | 0.49 ± 0.04 |
| ProTαN | 0.10 ± 0.05 | 3.17 ± 0.02 | 0.5 ± 0.1 |
| ProTαC | 0.08 ± 0.03 | 4.02 ± 0.01 | 0.4 ± 0.1 |

**Table S5. Urea concentration dependence obtained with Eq. S3 (Sanchez model, Fig. S1); parameters from fits with binding model (Eq. 2)**

| | $K$ | $R_{g0}$ (nm) | $\rho$ |
|---|---|---|---|
| Csp*Tm* | 0.20 ± 0.02 | 1.50 ± 0.01 | 1.02 ± 0.07 |
| IN | 0.18 ± 0.03 | 2.41 ± 0.01 | 0.32 ± 0.02 |
| ProTαN | 0.14 ± 0.07 | 3.17 ± 0.02 | 0.25 ± 0.06 |
| ProTαC | 0.10 ± 0.05 | 3.82 ± 0.01 | 0.29 ± 0.08 |

**Table S6. GdmCl concentration dependence obtained with Eq. S1 (Gaussian chain, Fig. 4 and Fig. S1); parameters from fits with polyelectrolyte theory (Eq. S8)**

| | $K$ | $\nu b^3$ ($nm^3$) | $\rho$ |
|---|---|---|---|
| Csp*Tm* | 2 ± 2 * | 0.05 ± 0.01 | 0.8 ± 0.4 |
| IN | 0.39 ± 0.03 | 0.36 ± 0.01 | 0.36 ± 0.01 |
| ProTαN | 0.7 ± 0.2 | 0.86 ± 0.05 | 0.31 ± 0.02 |
| ProTαC | 0.5 ± 0.2 | 0.72 ± 0.08 | 0.34 ± 0.03 |

*The large error reflects the inadequacy of the model for Csp*Tm* with a constant value of the excluded volume (see Fig. S3).

# 3. Probing Protein-Chaperone Interactions with Single-Molecule Fluorescence Spectroscopy

*Angew. Chem. Int. Ed.*, **2008**, 47, 6184 –6188

**Protein Folding**

DOI: 10.1002/anie.200800298

# Probing Protein–Chaperone Interactions with Single-Molecule Fluorescence Spectroscopy**

*Frank Hillger, Dominik Hänni, Daniel Nettels, Sonja Geister, Michelle Grandin, Marcus Textor, and Benjamin Schuler**

Molecular chaperones are an essential part of the cellular machinery that aids protein folding and assembly in vivo. Particularly remarkable are the members of the Hsp60 class, which encapsulate the folding protein in a central, closed cavity; the most well-studied example is the bacterial GroEL/ES system. Work of the past two decades has resolved many aspects of the processes involved.[1] However, remarkably little is known about the influence of the chaperone on the conformational distributions and folding mechanisms of its substrate proteins.[2] Because of the structural heterogeneity of the nonnative substrate bound to a molecular machine in the $10^6$ Da range, its experimental investigation has been difficult with established ensemble methods.[2] Since single-molecule spectroscopy, in particular in combination with Förster resonance energy transfer (FRET), can provide distance and orientational information free of ensemble averaging[3] and allows intramolecular distance dynamics to be observed at equilibrium,[4,5] it is a promising approach to address such questions.[6] Herein, we show how single molecule FRET can be utilized to investigate the nonnative conformation and dynamics of bovine rhodanese, a classic chaperone substrate protein,[7,8] upon interaction with GroEL.

To obtain a transfer efficiency signature suitable for discriminating native and nonnative conformations, two rhodanese variants with complementary donor and acceptor positions (Figure 1) were investigated. Figure 1 c–j shows the transfer efficiency histograms determined from photon bursts originating from individual labeled rhodanese molecules freely diffusing through the observation volume of the confocal instrument. As expected, the rhodanese variant with the labels at the domain interface (Figure 1 a) shows a mean transfer efficiency $\langle E \rangle$ close to 1 in its native state (Figure 1 c); for the variant with labels at the ends of the interdomain linker (Figure 1 b), $\langle E \rangle = 0.69$ (Figure 1 d), which corresponds to a distance of 4.7 nm, in good agreement with the distance of 4.5 nm in the crystal structure.[9] In the unfolded state at 5 M guanidinium chloride (GdmCl), $\langle E \rangle$ scales with the sequence separation of the labeling sites (Figure 1 e,f), as expected.

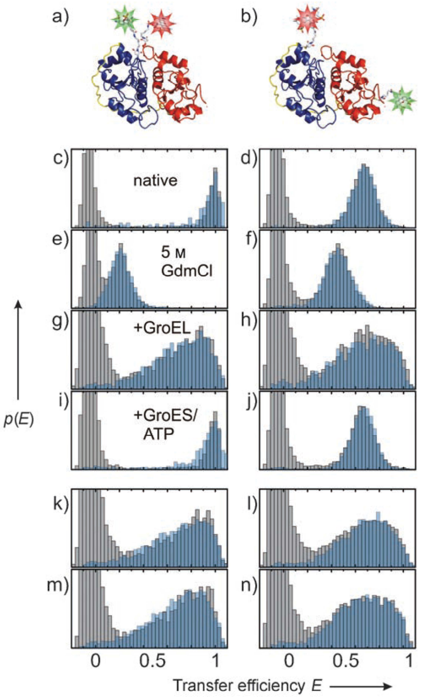

***Figure 1.*** Native structures (based on PDB ID1RHD) and transfer efficiency ($E$) histograms of rhodanese variants D102C–D219C (interface variant, panels on the left) and K135C–K174C (linker variant, panels on the right). a,b) Alexa Fluor488 and Alexa Fluor594 were coupled to the cysteine residues introduced by site-directed mutagenesis. c–j) $E$ histograms of rhodanese variants under native (c,d) and denaturing conditions (5 M GdmCl, e,f), bound to GroEL upon dilution from GdmCl (g,h), and after refolding by addition of GroES/ATP (i,j), bound to GroEL after incubation with folded rhodanese at 30 °C for 16 h (k,l), and after dilution from 0.1 M phosphoric acid (m,n).The gray histograms were recorded with donor excitation only. For the blue histograms, pulsed interleaved excitation[14] was used. $p(E)$ = relative event frequency.

---

[*] Dr. F. Hillger, D. Hänni, Dr. D. Nettels, S. Geister, Prof. Dr. B. Schuler
Universität Zürich, Biochemisches Institut
8057 Zürich (Switzerland)
Fax: (+41) 44-635-5907
E-mail: schuler@bioc.uzh.ch
Homepage: http://www.bioc.uzh.ch/schuler

Dr. M. Grandin, Prof. Dr. M. Textor
ETH Zürich, Departement Materialwissenschaft
8093 Zürich (Switzerland)

[**] We thank A. Szabo for discussions and advice on anisotropy simulations, A. Plückthun and M. Kawe for discussions and samples of GroEL, H. Hofmann and D. Streich for discussions, the late P. Horowitz for a plasmid encoding rhodanese, and G. Lorimer for a plasmid encoding SR1. This work was supported by the VolkswagenStiftung, the Schweizerische Nationalfonds, the Swiss National Center of Competence in Research for Structural Biology, and the Human Frontier Science Program.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200800298.

Upon dilution of labeled rhodanese unfolded in GdmCl into buffer containing an excess of unlabeled GroEL, rhodanese becomes bound to the chaperone quantitatively, as evidenced by analytical size-exclusion chromatography (data not shown). We exclude the possibility of substrate protein binding to both chaperone rings by using the single-ring variant of GroEL, SR1,which binds to the substrate in a 1:1 complex.[10,11] Experiments with wild-type tetradecameric GroEL gave results essentially identical to the ones presented here. The transfer efficiency histograms of SR1-bound rhodanese (Figure 1 g,h) exhibit a pronounced broadening, indicating the presence of static heterogeneity on the observation time scale ($\approx 1$ ms, duration of a fluorescence burst). For a random conformational distribution, we would expect a transfer efficiency that scales with the sequence separation of the dyes, as in the denaturant-unfolded state (Figure 1 e,f). In contrast, we observe maxima of the transfer efficiency histograms close to the values found in the native state (Figure 1 c,d), suggestive of a bias towards the native topology for rhodanese bound by the chaperone. The presence of very low intramolecular transfer efficiencies that could be hidden under the "donor only" peak[12] at $E \approx 0$ was excluded in experiments using alternating excitation of donor and acceptor[13,14] (Figure 1 c–n). The slight but reproducible difference in shape between the transfer efficiency histograms of the two chaperone-bound rhodanese variants (Figure 1 g,h) suggests that the $E$ histograms provide a characteristic signature for the conformation of the substrate protein. Remarkably, the shapes of the transfer efficiency histograms are independent of how rhodanese is denatured (Figure 1 g,h,k–n), implying that the chaperone-bound conformation does not reflect the conformational distribution under unfolding conditions, but rather resembles a folding intermediate that is formed rapidly upon dilution into the SR1 solution (Figure 1 g,h,m,n), and that is also accessible from the native state under mildly destabilizing conditions (Figure 1 k,l).[15] After addition of ATP and the cochaperone GroES to the rhodanese–GroEL complex, the $E$ distributions characteristic of the native structures are recovered[*] (Figure 1 c,d,i,j), demonstrating that labeled rhodanese is a fully functional chaperone substrate.

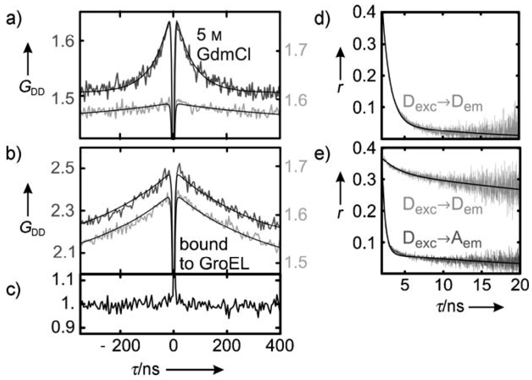

***Figure 2.*** Dynamics of rhodanese. a–c) Donor–donor fluorescence intensity autocorrelation functions $G_{DD}$ from Hanbury Brown and Twiss start–stop experiments.[4] Correlation functions are shown for the linker variant unfolded in 5 M GdmCl (a) and bound to GroEL (b). Dark gray lines show the correlation functions for the FRET-labeled, light gray lines for the donor-only-labeled rhodanese K174C. The black curves in (a) and (b) show fits to the correlation data including photon antibunching.[4] c) The normalized ratio of the two correlation functions from (b) indicates the absence of distance dynamics. d,e) Anisotropy decays for donor-only-labeled rhodanese K174C (light gray) under denaturing conditions (d) and bound to GroEL (e). The dark gray data in (e) show the fluorescence anisotropy decay of the acceptor upon excitation of the donor for the linker variant bound to GroEL. The black curves in (d) and (e) represent fits to Equation (1).

To probe the dynamics of the rhodanese–chaperone complex, we used correlation experiments employing a Hanbury Brown and Twiss setup.[4] Figure 2 a shows that rhodanese unfolded in 5 M GdmCl exhibits rapid intramolecular chain dynamics on a time scale of $\approx 70$ ns. This time scale is very similar to that observed for the unfolded cold shock protein Csp*Tm*[4] and the Sup35 NM domain.[**][16] How do the dynamics of the denatured state change upon association with GroEL? The same measurement on rhodanese bound to GroEL yields a correlation with a decay of 0.2 µs (Figure 2 b), which at first sight could be misinterpreted as slowed distance dynamics. But the pronounced sensitivity of the correlation amplitude on the directions of polarization that are correlated (Figure S1 in the Supporting Information) indicates a strong contribution from rotational motion of the entire GroEL–rhodanese complex, which occurs exactly on this time scale.[*][18] To quantify the relative contributions of rotational and distance dynamics, we compare GroEL-bound rhodanese labeled with a FRET pair to GroEL-bound rhodanese labeled only with a donor chromophore. As shown in Figure 2 b, the two samples exhibit the same decay time of the correlation function. The ratio of the two correlations does not indicate the presence of an additional component (Figure 2 c), suggesting that the observed correlation is entirely due to rotation, and that distance dynamics are absent on this time scale. Additional evidence for the lack of distance dynamics comes from the pronounced intensity correlations of polarized acceptor emission upon donor excitation, which exhibit the same 0.2 µs decay (Figure S2 in the Supporting Information). This result shows that the relative orientation of donor and acceptor is rather invariant on this time scale, arguing that the same is true for their distance. The nanosecond chain

---

[*] Under our conditions, folded rhodanese is not confined within the cage.

[**] Note, however, that (in contrast to Csp*Tm*[4]), even singly labeled rhodanese exhibits some bunching on this time scale, albeit with lower amplitude (Figure 2 a), which complicates a quantitative analysis. This behavior is similar to recent observations for a Sup35 fragment, which were attributed to quenching of the fluorophores by aromatic residues in the chain.[16]

[*] In contrast to the magic angle configuration possible in conventional fluorimeters, the geometry of confocal epifluorescence instruments complicates the elimination of polarization effects on the correlation functions.[17]

dynamics observed in denaturant-unfolded rhodanese are thus suppressed when the protein is bound to the chaperone.

To investigate the presence of distance dynamics on longer time scales, we first employed subpopulation-specific fluorescence correlation spectroscopy on freely diffusing rhodanese–GroEL complexes (Figure 3). We correlated only signal from FRET-labeled molecules with $E > 0.2$ to minimize the contribution from donor-only labeled species, and we used donor–acceptor cross-correlation analysis to minimize the contribution of microsecond triplet dynamics.[19] Distance fluctuations would then result in an anticorrelated signal, that is, a rise in the correlation function. However, the correlation curves show no evidence for the presence of distance fluctuations up to $\approx 100$ µs. To extend the accessible time scales beyond the diffusion time through the confocal volume, SR1–rhodanese complexes were immobilized on cover slides coated with biotinylated poly(L-lysine)-*graft*-poly(ethylene glycol) (PLL-*g*-PEG).[20] Individual complexes on the surface were identified by sample scanning and were observed individually for several seconds until the chromophores bleached. Surprisingly, donor–acceptor cross-correlation analysis of these data indicates the absence of long-range distance dynamics in chaperone-bound rhodanese even on long time scales. The decay of the correlation function setting in at times $> 10$ ms is caused by irreversible photobleaching, as indicated by the divergence of donor–acceptor and acceptor–donor cross-correlations (Figure 3).[21] The absence of large-amplitude distance fluctuations is also supported by the large width of transfer efficiency histograms from different observation or binning times (Figure 3, insets), indicating the presence of static heterogeneity on time scales up to at least 100 ms.

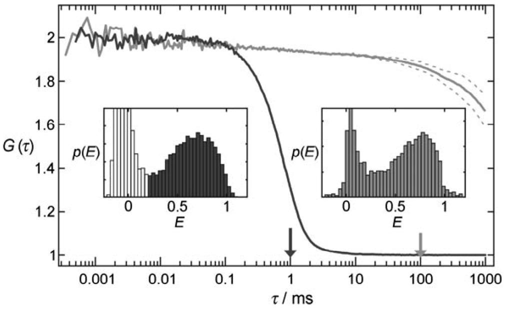

***Figure 3.*** Normalized fluorescence intensity donor–acceptor cross-correlations and transfer efficiency histograms of the rhodanese-linker variant bound to GroEL (free diffusion: dark gray, surface-immobilized: light gray; corresponding binning times for $E$ histograms indicated by arrows). Dashed lines indicate the individual cross-correlations (D→A and A→D, shown for $\tau > 0.5$ ms). For freely diffusing molecules, only events with $E > 0.2$ were used for the correlation (dark gray).

Finally, we need to establish the structural origin of the large width of the transfer efficiency distributions of chaperone-bound rhodanese (Figure 1). Static heterogeneity of the transfer rate can originate from either a distribution of intramolecular distances or a distribution of donor–acceptor orientations.[*] For rhodanese singly labeled with donor or acceptor and unfolded in 5 M GdmCl, the anisotropy decay $r(t)$ is dominated by a single component with a time constant of $\approx 1$ ns (Figure 2 d), indicating rapid and complete reorientation of the dyes, and thus justifying the common approximation of $\kappa^2 \approx 2/3$ for the orientational factor in Förster theory.[22] Upon binding to SR1, however, the anisotropy of all singly labeled variants (D102C, K135C, K174C, D219C) increases drastically, and the majority of the anisotropy decay occurs on the time scale of rotation of the entire rhodanese–SR1 complex ($> 100$ ns, Figure 2 e). Consequently, the orientational restriction of the dyes must be taken into account to obtain distance information.

To this end, we analyze the fluorescence anisotropy decays of our singly labeled rhodanese variants with Equation (1), which describes the decay as the combined effect of restricted dye rotation ($\tau_{\mathrm{eff}}$) and the rotational motion of the entire protein–chaperone complex ($\tau_{\mathrm{M}}$).[23]

$$r(t) = \left((r_0 - r_\infty)\mathrm{e}^{-t/\tau_{\mathrm{eff}}} + r_\infty\right)\mathrm{e}^{-t/\tau_{\mathrm{eff}}} \qquad (1)$$

Here, $r_0$ is the limiting anisotropy of the dyes,[**] and $r_\infty$ is the residual anisotropy assuming no rotation of the macromolecule carrying the dye. Assuming restricted angular diffusion in a cone as the simplest plausible model for the motion of the chromophores[24] (Figure 4 a), the semiangle $\Theta_{\max}$ of the cone can be calculated from Equation (2),[23,25] yielding for all our variants and dyes values between 17° and 19°.[***]

$$r_\infty = r_0\left(\frac{1}{2}\cos\Theta_{\max}(1 + \cos\Theta_{\max})\right)^2 \qquad (2)$$

Important additional information about the relative orientation of the dyes comes from the anisotropy decay of the acceptor upon donor excitation (Figure 2 e): in this case, the residual anisotropy approaches zero for both chaperone-bound variants, indicating an angular distribution of the cone axes that is close to random. An alternative explanation, a narrow relative orientation close to the magic angle of 54.7°, can be excluded, because this would result in an apparent fundamental anisotropy of zero for the acceptor anisotropy decay upon donor excitation, which is incompatible with our obervations (Figure 2 e). Additionally, a narrow distribution

---

[*] Heterogeneity in the quantum yields of the dyes originating, for example, from differences in the local environment can be excluded because of the agreement of fluorescence lifetimes (both in ensemble and single-molecule measurements) of the acceptor in FRET-labeled and the donor in singly labeled rhodanese on GroEL, respectively, with the lifetimes of the dyes on protein unfolded in 5 M GdmCl.

[**] $r_0 = 0.38$ was determined in a matrix of 99 % glycerol at −10 °C.

[***] The lack of binding to GroEL of free dyes and several other small proteins and peptides labeled with the same dyes (size exclusion chromatography data not shown) indicates that the interaction of rhodanese with GroEL is dominated by the polypeptide and that the orientational restriction of the dyes results largely from steric constraints in the rhodanese–chaperone complex.

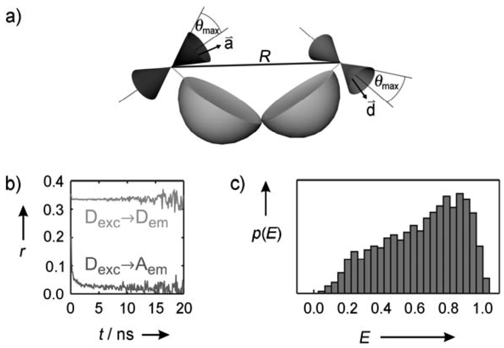

***Figure 4.*** Simulations and structural interpretation. a) Assuming angular diffusion of the fluorophore dipoles $\vec{d}$ and $\vec{a}$ in cones of half angle $\Theta_{max}$ on the protein surface, and a narrow Gaussian distribution of distance $R$, we can account for both the characteristic anisotropy decays[34] (b) and the broad transfer efficiency histograms (c) observed experimentally (cf. Figures 1 and 2e). The two domains of rhodanese are indicated by the gray hemispheres in (a).

of relative orientations would at the same time require a broad distribution of distances to account for the broad transfer efficiency histograms we observe (Figure 1g,h,k–n), but this combination is physically implausible.

To interpret the experimental results quantitatively, we thus simulated the transfer process between orientationally restricted dipoles based on the simplest plausible model for our system (Figure 4): we assume that the relative orientation of the cones is fixed for every individual rhodanese–GroEL complex but randomly distributed from molecule to molecule. This assumption leads to anisotropy decays (Figure 4b) very similar to the experimental ones (Figure 2e), even for the characteristic decay of the acceptor anisotropy upon donor excitation. If we now assume a normal distribution of interdye distances $R$ (Figure 4a) and adjust its mean and standard deviation to maximize the agreement between simulated and observed transfer efficiency histograms (Figure 4c and Figure S3 in the Supporting Information), we obtain distance distributions for GroEL-bound rhodanese with a mean distance of $\approx 4.5 \pm 0.5$ nm and a width of $\approx 0.5 \pm 0.2$ nm for both variants. For the linker variant, this is approximately the same value as in the native structure, but for the interface variant it is significantly larger, suggesting a large separation of the two rhodanese domains. If we used the mean distances for our two variants as constraints to adjust the relative orientation of the native domains, we would obtain a rhodanese conformation that is highly suggestive of binding to the rim of the GroEL ring, which is lined by hydrophobic residues that act as binding sites,[26] an arrangement that is in accord with a number of previous results.[27–29]

In summary, we have used a novel analysis combining time-resolved fluorescence anisotropy decays, single-molecule FRET experiments, and simulations to obtain quantitative information from a system with orientationally restricted chromophores, a situation that has been observed repeatedly in FRET experiments involving protein–chaperone interactions.[6,30] In our analyis, the donor and acceptor anisotropy decays define the opening angles of the cones constraining fluorophore rotation; the acceptor decay upon donor excitation constraines the relative orientational distribution of the cones; and the shape of the transfer efficiency histograms then define the mean and width of the distance distributions. What emerges from these measurements, together with the long-range dynamic information available from subpopulation-specific correlation functions, is the picture of a rather well-defined ensemble of rhodanese conformations that resembles a partially structured folding intermediate when bound to the chaperone GroEL. Interestingly, the lack of long-range distance dynamics does not seem to preclude local structural fluctuations evident from protease susceptibility,[27,28] or NMR[18,31] or fluorescence spectroscopy.[8,27] Our results illustrate the potential for extracting quantitative structural information from FRET experiments even in cases where large anisotropies demand that orientational effects be taken into account, and provide an important step towards investigating the role of cellular factors in protein folding.

## *Experimental Section*

Proteins were prepared as described previously.[10,32] Binding of rhodanese to SR1 was achieved as follows: A) Rhodanese unfolded in 5 M GdmCl was rapidly diluted tenfold into folding buffer (0.1 M potassium phosphate, 5 mM magnesium chloride, 200 mM 2-mercaptoethanol, 0.001 % Tween 20, 1 mM EDTA, pH 7.0) containing at least a tenfold molar excess of SR1 heptamers. B) Same as (A), but unfolding was carried out in 0.1 M phosphoric acid. C) Rhodanese was incubated at 30 °C for 16 h in folding buffer with a tenfold molar excess of SR1 heptamers. Complete binding was assessed on a TSK 5000 PWXL column (TOSOH Bioscience) with fluorescence detection.

For surface immobilization, SR1 was biotinylated using (+)-biotin *N*-hydroxysuccinimide ester in a molar ratio of 1:7. A solution of 0.1 mg mL$^{-1}$ PLL(20)-g[3.5]-PEG(2)/PEG(3.4)-biotin (50 %)[20] in 10 mM potassium phosphate, pH 7.0, was applied to a custom-made quartz flow cell. After 15 min of incubation, the flow cell was washed with 0.1 M potassium phosphate, 5 mM magnesium chloride, 1 mM EDTA, pH 7.0; then 1 mg mL$^{-1}$ avidin was applied in the same buffer. After 15 min of incubation, the flow cell was washed thoroughly, and 250–500 nM GroEL–rhodanese preparation was applied. The cell was incubated for 5 min and then washed with buffer.

Anisotropy decay data were recorded with a custom-built fluorescence lifetime spectrometer using 1 μM samples of labeled protein. Single-molecule FRET measurements were performed as previously described[4,32,33] using an adapted MicroTime 200 confocal microscope (PicoQuant, Berlin). For additional details, see the Supporting Information.

Received: January 21, 2008
Revised: April 29, 2008

**Keywords:** chaperones · correlation spectroscopy · FRET · protein folding · single-molecule studies

---

[1] D. Thirumalai, G. H. Lorimer, *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 245–269; H. Grallert, J. Buchner, *J. Struct. Biol.* **2001**, *135*, 95–103; J. C. Young, V. R. Agashe, K. Siegers, F. U. Hartl, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 781–791; A. L. Horwich, G. W. Farr, W. A. Fenton, *Chem. Rev.* **2006**, *106*, 1917–1930.

[2] W. A. Fenton, A. L. Horwich, *Q. Rev. Biophys.* **2003**, *36*, 229–256.

[3] G. Haran, *J. Phys. Condens. Matter* **2003**, *15*, R1291–R1317; B. Schuler, *ChemPhysChem* **2005**, *6*, 1206–1220; X. Michalet, S. Weiss, M. Jäger, *Chem. Rev.* **2006**, *106*, 1785–1813; B. Schuler, W. A. Eaton, *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

[4] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655–2660.

[5] E. V. Kuzmenkina, C. D. Heyes, G. U. Nienhaus, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15471–15476.

[6] S. Sharma, K. Chakraborty, B. K. Müller, N. Astola, Y. C. Tang, D. C. Lamb, M. Hayer-Hartl, F. U. Hartl, *Cell* **2008**, *133*, 142–153.

[7] J. A. Mendoza, E. Rogers, G. H. Lorimer, P. M. Horowitz, *J. Biol. Chem.* **1991**, *266*, 13044–13049.

[8] J. Martin, T. Langer, R. Boteva, A. Schramel, A. L. Horwich, F. U. Hartl, *Nature* **1991**, *352*, 36–42.

[9] J. H. Ploegman, G. Drent, K. H. Kalk, W. G. J. Hol, *J. Mol. Biol.* **1978**, *123*, 557–594.

[10] A. L. Horwich, S. G. Burston, H. S. Rye, J. S. Weissman, W. A. Fenton, *Methods Enzymol.* **1998**, *290*, 141–146.

[11] J. S. Weissman, H. S. Rye, W. A. Fenton, J. M. Beechem, A. L. Horwich, *Cell* **1996**, *84*, 481–490.

[12] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2754–2759.

[13] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8936–8941.

[14] B. K. Müller, E. Zaychikov, C. Bräuchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522.

[15] T. Shibatani, G. Kramer, B. Hardesty, P. M. Horowitz, *J. Biol. Chem.* **1999**, *274*, 33795–33799; K. E. Smith, P. A. Voziyan, M. T. Fisher, *J. Biol. Chem.* **1998**, *273*, 28677–28681.

[16] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2649–2654.

[17] Ü. Mets in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer, Berlin, **2001**.

[18] R. Horst, E. B. Bertelsen, J. Fiaux, G. Wider, A. L. Horwich, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12748–12753.

[19] P. Schwille in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer, Berlin, **2001**; C. Eggeling, P. Kask, D. Winkler, S. Jäger, *Biophys. J.* **2005**, *89*, 605–618.

[20] N. P. Huang, J. Vörös, S. M. De Paul, M. Textor, N. D. Spencer, *Langmuir* **2002**, *18*, 220–230.

[21] C. Eggeling, J. Widengren, L. Brand, J. Schaffer, S. Felekyan, C. A. M. Seidel, *J. Phys. Chem. A* **2006**, *110*, 2979–2995.

[22] B. W. Van Der Meer, G. Coker III, S. Y. S. Chen, *Resonance Energy Transfer: Theory and Data*, VCH, New York, **1994**.

[23] G. Lipari, A. Szabo, *Biophys. J.* **1980**, *30*, 489–506.

[24] G. F. Schröder, U. Alexiev, H. Grubmüller, *Biophys. J.* **2005**, *89*, 3757–3770.

[25] K. Kinosita, S. Kawato, A. Ikegami, *Biophys. J.* **1977**, *20*, 289–305.

[26] K. Braig, Z. Otwinowski, R. Hegde, D. C. Boisvert, A. Joachimiak, A. L. Horwich, P. B. Sigler, *Nature* **1994**, *371*, 578–586.

[27] J. A. Mendoza, M. C. Butler, P. M. Horowitz, *J. Biol. Chem.* **1992**, *267*, 24648–24654.

[28] R. Hlodan, P. Tempst, F. U. Hartl, *Nat. Struct. Biol.* **1995**, *2*, 587–595.

[29] P. Thiyagarajan, S. J. Henderson, A. Joachimiak, *Structure* **1996**, *4*, 79–88; A. van der Vaart, J. P. Ma, M. Karplus, *Biophys. J.* **2004**, *87*, 562–573; G. W. Farr, K. Furtak, M. B. Rowland, N. A. Ranson, H. R. Saibil, T. Kirchhausen, A. L. Horwich, *Cell* **2000**, *100*, 561–573.

[30] Z. Lin, H. S. Rye, *Mol. Cell* **2004**, *16*, 23–34.

[31] R. Zahn, S. Perrett, G. Stenberg, A. R. Fersht, *Science* **1996**, *271*, 642–645; R. Zahn, C. Spitzfaden, M. Ottiger, K. Wüthrich, A. Plückthun, *Nature* **1994**, *368*, 261–265.

[32] F. Hillger, D. Nettels, S. Dorsch, B. Schuler, *J. Fluoresc.* **2007**, *17*, 759–765.

[33] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 105–110.

[34] Rotation of the rhodanese-GroEL complex was not included in the simulations, which therefore lack the slow relaxation component ($\tau_M = \infty$ in Eq. 1).

# Supporting information

## Probing protein-chaperone interactions with single molecule fluorescence spectroscopy

Frank Hillger, Dominik Hänni, Daniel Nettels, Sonja Geister, Michelle Grandin, Marcus Textor, and Benjamin Schuler

### Materials and Methods

***Protein Preparation and Labeling.*** Cysteine variants of rhodanese were prepared and labeled with Alexa 488 and Alexa 594 ($R_0$ = 5.4 nm) as described previously[1]. Cysteine residues were introduced by site-directed mutagenesis at positions 135 and 174 ("linker variant") and 102 and 219 ("interface variant"), respectively. After labeling, thiosulfate was removed from the active site of rhodanese by reaction with potassium cyanide and subsequent size exclusion chromatography. GroEL-SR1 was expressed and purified essentially as described by Horwich *et al.*[2] and stored at -80 °C as a suspension of precipitate in 2.7 M ammonium sulfate.

***Preparation of Rhodanese-Chaperone Complexes.*** SR1 ammonium sulfate precipitate was resolubilized at 10 to 20 µM SR1 (heptamer) in folding buffer (0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 200 mM 2-mercaptoethanol (Fluka Ultra), 0.001 % Tween 20 (Pierce), 1 mM EDTA (Biosolve), pH 7.0). Binding of rhodanese to SR1 was achieved as follows: **A**) Rhodanese unfolded in 5 M GdmCl (Pierce), 0.1 M potassium phosphate, 200 mM 2-mercaptoethanol, 0.001 % Tween 20, 1 mM EDTA, pH 7.0 was rapidly diluted tenfold into folding buffer containing at least a tenfold molar excess of SR1 heptamers. **B**) As **A,** but unfolding was carried out by tenfold dilution of native rhodanese into 0.1 M phosphoric acid. **C**) Rhodanese was incubated at 30 °C over night in folding buffer containing a 10-fold molar excess of SR1 heptamers. Complete binding was assessed by analytical size exclusion chromatography on a TSK 5000 PWXL (TOSOH Bioscience) and fluorescence detection. In all cases, rhodanese was bound completely to SR1. Single-molecule experiments and size exclusion chromatography indicated the absence of unbound rhodanese molecules or aggregates[1].

Note that rhodanese shows a pronounced collapse upon dilution from denaturant within the dead time of ensemble stopped-flow mixing experiments, indicating the formation of a compact structural ensemble before the slower phases of folding ensue. The dominant phase of refolding for our labeled rhodanese variants occurs with half times of ~15 min, within the range of reactivation times of unlabeled rhodanese reported in the literature[3], suggesting that labeling does not strongly interfere with the folding mechanism. The widths of the transfer efficiency histograms (Fig. 1 main text) for native, unfolded, and refolded rhodanese interface variant were 0.08±0.01, 0.08±0.01, and 0.10±0.01, respectively, and for the linker variant 0.11±0.01, 0.10±0.01, and 0.10±0.01, respectively. At the exceedingly low protein concentrations used in our experiments, aggregation rates become negligible and the native state is reached reversibly upon refolding also in the absence of GroEL/ES.

***Biotinylation and Surface Immobilization of GroEL/SR1.*** Ammonium sulfate-precipitated SR1 was transferred to 100 mM $NaHCO_3$ (Roth) using a 5 ml HiTrap desalting column (GE Amersham), followed by concentration to 30-40 mg/ml in Centricon YM-3 filter devices (Amicon). (+)-Biotin N-hydroxy-succinimide ester (Sigma) was dissolved in dry DMSO (Pierce) and added to concentrated SR1 in a molar ratio of 1 to 7. After incubation for 4 hours, the reaction mixture was purified and transferred to 0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 1 mM EDTA (Biosolve), pH 7.0, using a Superdex 200 10/300 GL (Amersham) gel filtration column. Biotinylation was analyzed using ESI TOF mass spectroscopy and samples were stored at 5 mg/ml at -80 °C. 0.1 mg/ml PLL(20)-g[3.5]-PEG(2)/PEG(3.4)-Biotin (50%)[4] was dissolved in 10 mM potassium phosphate (Fluka Ultra), pH 7.0, and applied to a custom-made quartz flow cell. After 15 min of incubation, the flow cell was washed with 0.1 M potassium phosphate (Fluka Ultra), 5 mM magnesium chloride (Roth), 1 mM EDTA (Biosolve), pH 7.0, then 1 mg/ml Avidin D (Vector) were applied in the same buffer. After 15 min of incubation, the flow cell was washed vigorously. The GroEL/rhodanese complex was prepared as described above and separated from monomeric rhodanese and rhodanese aggregates by size exclusion chromatography. 250-500 nM of GroEL/rhodanese preparation were applied to the flow cell and incubated for 5 min before the cell was washed with buffer.

***Single Molecule Fluorescence Spectroscopy.*** Observations of single-molecule fluorescence were made using a MicroTime 200 confocal single molecule instrument (PicoQuant, Berlin, Germany) equipped with a Olympus UplanApo 60x/1.20W objective and four detection channels. Donor excitation pulses of 470 nm wavelength are generated by a diode laser (LDH 470, PicoQuant). Except for PIE measurements (see below) the pulse repetition rate was 40 MHz, resulting in an average power of 100 μW entering the completely illuminated back aperture of the objective. Laser light and fluorescence emission are separated using a dual band beam splitter (z488/568, Chroma), followed by a 100 μm pinhole. Sample fluorescence was further separated by a polarizing beam splitter cube into components parallel and perpendicular polarized with respect to the excitation light. Subsequently, both components were further divided into donor and acceptor photons by means of dichroic mirrors (585DCXR, Chroma), filtered (donor emission filters: Chroma HQ525/50, Omega 525AF45, acceptor emission filters: Chroma HQ650/100 and HQ600LP), and focused on avalanche photodiodes (PerkinElmer Optoelectronics SPCM-AQR-15). A TimeHarp 200 counting card (PicoQuant) records the arrival time of every detected photon with a time resolution of 100 ns. In addition, the time separation between excitation pulse and photon detection is stored with 38 ps resolution. For pulsed interleaved excitation[5] PIE, the wavelength range used for acceptor excitation was selected with a z582/15 band pass filter (Chroma) from the emission of an SC-450-4 supercontinuum fiber laser (Fianium, UK) driven at 15 MHz. The emission is monitored by a fast photodiode (DET10A/M, Thorlabs), which triggers the LDH 470 driver to deliver single laser pulses. The time separation between the acceptor and donor excitation pulses of 33 ns was adjusted with a delay line. FCS data down to 0.1 μs lag time (Fig. 3) were recorded and calculated using the software SymphoTime (PicoQuant). Fluorescence correlation data in the sub-microsecond range (down to 256 ps lag time, Fig. 2) were recorded and analyzed as described previously using a Hanbury Brown and Twiss detection scheme[6], a continuous wave solid-state diode-pumped laser (Coherent Sapphire 488-200) operating at 488 nm (average radiant power at the sample: 100μW) and a PicoHarp 300 or a HydraHarp 400 photon counting module (PicoQuant). Subpopulation-specific donor-acceptor cross-correlations were obtained by correlating only photons from bursts assigned to the unfolded subpopulation.

For experiments on surface-immobilized molecules, the MicroTime 200 (PicoQuant) was used in combination with a 488 nm cw laser. Sample scanning with an x/y-piezo-stage (P-733 and E-710, Physik Instrumente) in combination with custom-developed software allowed the automated identification and data acquisition of immobilized molecules. In area scans over regions of $80\times 80\,\mu\mathrm{m}^2$ (0.156 nm point resolution, 5 μW average radiant power at the sample), the locations of the individual molecules were identified and measured for five seconds each. For the data shown in Fig. 2 of the main text, 1297 trajectories were analyzed. To quantify the background intensity, 30 trajectories were measured at positions lacking labeled protein.

***Single Molecule Data Reduction and Analysis.*** Successive photons detected in any channel and separated by less than 100 μs were combined into one burst. A burst was retained as a significant event if the total number of counts exceeded 50 (after correction for background, differences in quantum yields, the different collection efficiencies of the detection channels, cross-talk, and direct excitation of the acceptor)[7]. In addition, bursts with a significant likelihood for acceptor photobleaching during the burst were discarded (see below). Transfer efficiencies were calculated as $E = n_A/(n_D + n_A)$, where $n_D$ and $n_A$ are the corrected donor and acceptor counts of each burst.

For calculating correlation functions from measurements on immobilized molecules, the trajectories were merged and donor and acceptor intensities cross-correlated. For normalization, each correlation was divided by the mean fluorescence calculated from the first 10 data points of each correlation. For transfer efficiency histograms, the trajectory data were binned in 100 ms intervals and corrected for background. For each trajectory the maximum change in the sum of donor and acceptor fluorescence signal was calculated and only trajectories were selected were this difference was >200 counts and did not exceed 1200 counts. The trajectories were truncated at 1 s to minimize the contribution of background. Transfer efficiency histograms were calculated for each bin of the trajectories using $E = n_A/(n_A + \gamma n_D)$, where $\gamma = 1.2$ corrects for the difference in quantum yield of donor and acceptor dyes as well as the detection efficiencies of the channels.

***Identification of Fluorescence Bursts Affected by Photobleaching.*** Assume that during a burst of duration $T$, $n_D$ donor photons and $n_A$ acceptor photons were detected at times $t_{D,1}\ldots t_{D,n_D}$ and $t_{A,1}\ldots t_{A,n_A}$, respectively. With the average values $\overline{t}_D = 1/n_D\sum_i t_{D,i}$ and $\overline{t}_A = 1/n_A\sum_i t_{A,i}$, we define the burst asymmetry as $\alpha_{DA} = \overline{t}_D - \overline{t}_A$. In the limit of $n_D, n_A \to \infty$, the asymmetry is zero if donor and acceptor dyes emit photons continuously with rates, $\tau_{D,A}^{-1} = n_{D,A}/T$, during the complete burst duration $T$. However, if the acceptor dye bleaches at a certain time during the photon burst, the acceptor photon rate will drop to the background level, and the donor rate will increase due to the interrupted energy transfer, resulting in a burst asymmetry $\alpha_{DA} > 0$. For finite photon numbers $n_D$ and $n_A$, shot noise of the calculated $\alpha_{DA}$ has to be taken into account. Here we calculate the standard deviation of $\alpha_{DA}$ for the ideal case that no photobleaching occurs during the burst. For a randomly chosen burst photon $i$, the detection time $t_i$ is then uniformly distributed, with a constant probability density function $p(t_i) = 1/T$, such that $\int_0^T p(t_i)dt_i = 1$. The expectation value for $t_i$ is then $\overline{t_i} = \int_0^T t_i\, p(t_i)dt_i = T/2$ and the variance is $\sigma_i^2 = \int_0^T (t_i - \overline{t_i})^2 p(t_i)dt_i = T^2/12$. With these relations, the variance of $\overline{t_D}$ and $\overline{t_A}$ defined above result as $\Delta\overline{t_{D,A}}^2 = \sum_i \sigma_i^2 / n_{D,A}^2 = T^2 \big/ 12 n_{D,A}$. Finally, we obtain for the standard deviation of $\alpha_{DA}$:

$$\sigma_{DA} = \frac{T}{2\sqrt{3}}\left(\frac{1}{n_D} + \frac{1}{n_A}\right)^{1/2}. \tag{1}$$

This approximation is valid if the dead time of the detectors ($\approx 100\,\mathrm{ns}$) is short compared to the inter-photon times $\tau_{D,A}$ during the bursts (on average ~10 µs). To exclude bursts with a significant deviation of $\alpha_{DA}$ from zero, indicating that a bleaching process is likely to have occurred, bursts with $|\alpha_{DA}| > \sigma_{DA}$ were not included in the analysis. An example of the analysis is shown in Fig. S4.

***Time-Resolved Anisotropy.*** Anisotropy decay data were recorded with a custom-built fluorescence lifetime spectrometer using samples of 1 µM labeled protein. The fluorophores were excited at 470 nm by vertically polarized laser pulses with 20 MHz repetition rate. Collimated fluorescence light, emitted in perpendicular direction with respect to the laser beam, passed an analyzing Glan-Thompson polarizer and a band pass filter (525AF45, Omega Filters, or HQ640/100, Chroma) before it was focused onto a microchannel plate photomultiplier tube (PMT R3809U-50, Hamamatsu). The output pulses of the PMT were pre-amplified (PAM 102-M, PicoQuant), and times between laser pulses and fluorescence photon arrival times were recorded by a photon counting module (PicoHarp 300, PicoQuant). Fluorescence decay histograms with 8 ps bin width were formed. The overall instrument response function (IRF) of the system was measured to have a FWHM of $\Delta_{IRF} = 80$ ps. Two fluorescence decays $I_V(t)$ and $I_H(t)$ were measured for each sample with the analyzing polarizer in vertical and in horizontal orientation, respectively. From these decays the anisotropy as a function of time was determined as

$$r(t) = \frac{I_V(t) - GI_H(t)}{I_V(t) + 2GI_H(t)},$$

where $G$ is a correction factor compensating for the slightly different instrumental detection sensitivity for vertically and horizontally polarized photons.

***Simulations.*** Resonance energy transfer in the presence of orientational constraints of donor and acceptor dyes was simulated using Mathematica (Wolfram Research) as follows. The orientational dynamics of the fluorophores were modeled as angular diffusion of the dye dipole in a cone[8], approximated by a random walk on the corresponding segment of a unit sphere[9]. For every simulated fluorescence burst, corresponding to an individual rhodanese-GroEL complex, we assumed a random, isotropic, but fixed, relative orientation of the cone axes for donor and acceptor, separated by a distance selected randomly from a normal distribution of given mean and standard deviation. The Brownian dynamics for the diffusive motion of the dye dipoles were combined with a Monte Carlo simulation of the photon emission process of the coupled dye pair[10]. After each time step of 10 ps, the transfer efficiency was calculated from the distance and the relative orientation of donor and acceptor dipoles[11]. For each emitted photon, the orientation of the donor dye at the time of absorption, the orientation of the emitting dye at the time of emission, and the time interval between donor excitation and photon emission were stored for the calculation of anisotropy decays. Each calculation consisted of 5000 simulated bursts of 50 photons each (corresponding to 5000 rhodanese-GroEL complexes with different relative cone orientations) to achieve similar statistics as in the experiment. From the complete data set, transfer efficiency histograms and anisotropy decays were calculated as shown in Figs. 4 and S3.

**Supporting Figures**

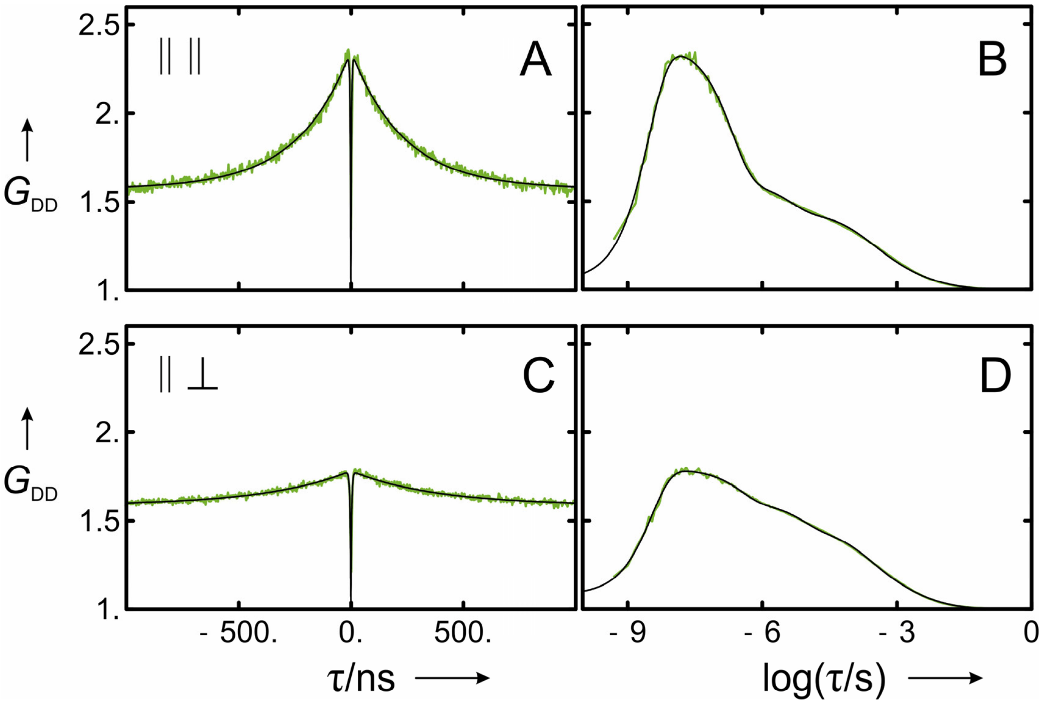

**Fig. S1.** *Dependence of the donor intensity autocorrelation function of GroEL-bound rhodanese on the polarizations used for signal correlation.* (A,B) Donor (horizontal polarization) – donor (horizontal polarization) intensity correlation of donor-only-labeled rhodanese bound to GroEL. (C,D) Donor (horizontal polarization) – donor (vertical polarization) intensity correlation. The pronounced difference in amplitude indicates that the correlation is caused by rotational motion of the entire rhodanese-chaperone complex[12]. The relative amplitudes are in good agreement with theoretical calculations of the polarization dependence[13] (not shown).

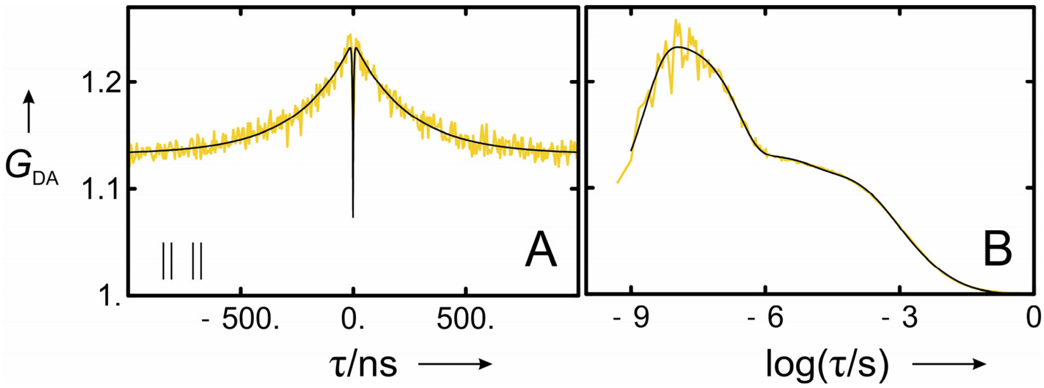

**Fig. S2.** *Donor (horizontal polarization)-acceptor (horizontal polarization) crosscorrelation of FRET-labeled chaperone-bound rhodanese.* The pronounced correlation on the timescale of rotational diffusion of the entire rhodanese-chaperone complex (A: linear, B: logarithmic scale) shows that the relative orientation of donor and acceptor chromophores in the chaperone-bound protein is rather fixed on the time scale of ~0.2 µs, such that a marked polarization component is retained in spite of energy transfer, implying that the intramolecular distance between the dyes also does not change on this time scale.

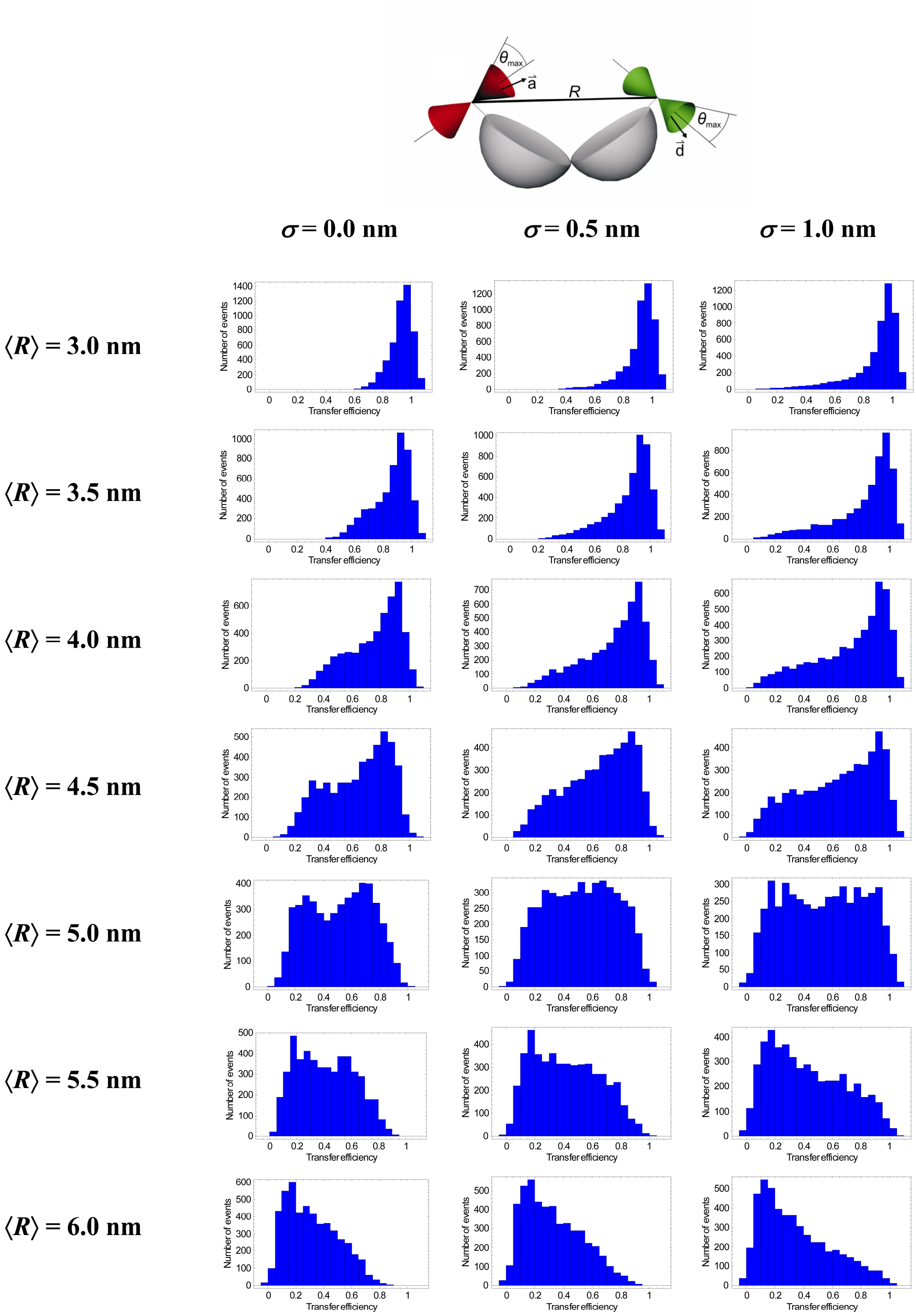

**Fig. S3.** *Simulated transfer efficiency histograms* with systematic variation of the distance distribution between donor and acceptor, assuming orientational restriction of the dyes as described in the main text and *Experimental Section*. Realistic background values were included for the calculation of the histograms. The shape of the histograms is very sensitive to the mean distance ⟨$R$⟩ and the width (standard deviation $\sigma$) of the distance distribution (here assumed to be Gaussian). Note in particular that for very short distances the transfer efficiency distribution is very narrow and close to $E = 1$. The Förster radius $R_0$ is 5.4 nm.

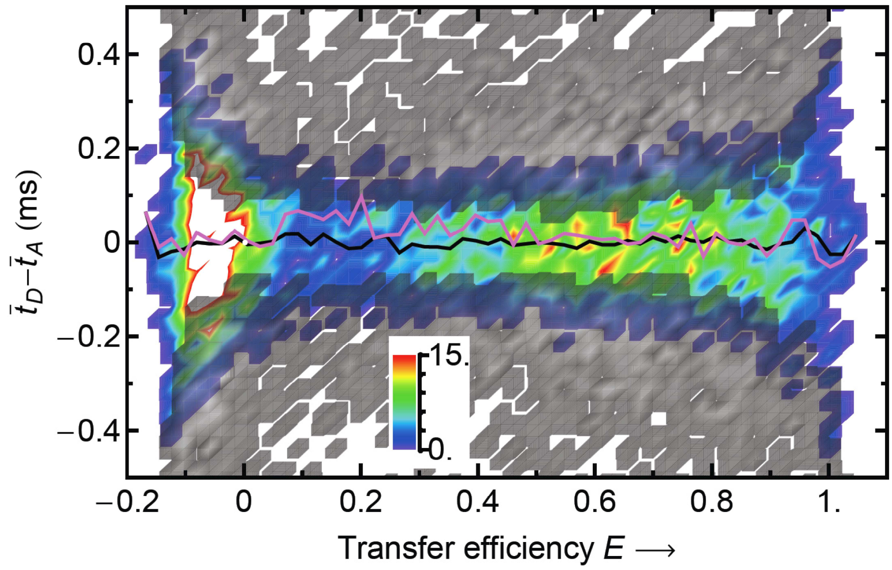

**Fig. S4.** *Identification of Fluorescence Bursts Affected by Photobleaching.* The burst asymmetry $\alpha$ is shown as a function of the FRET efficiency. The events with $|\alpha_{DA}| > \sigma_{DA}$ are shown in grey, the intensity scale for the remaining events is shown in color. The mean burst asymmetries before and after eliminating the asymmetric events are shown in purple and black, respectively, indicating exclusion of bursts where the acceptor was photobleached during the transit through the confocal volume.

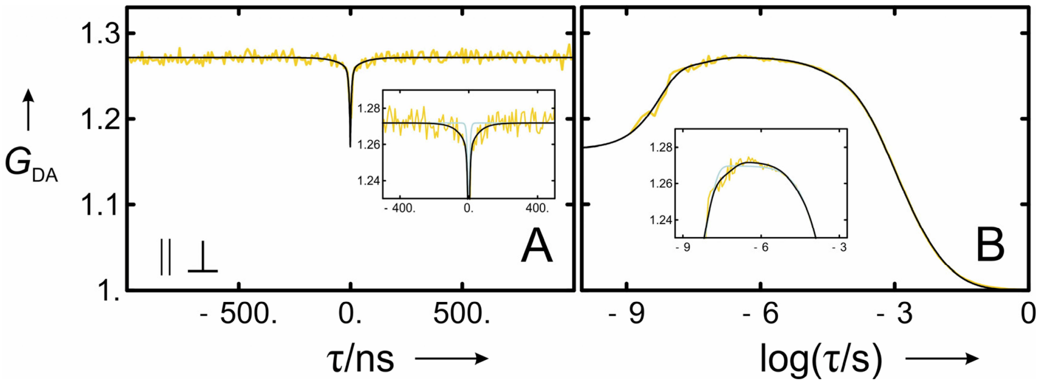

**Fig. S5.** *Anticorrelation of the Donor (horizontal polarization) - acceptor (vertical polarization) crosscorrelation caused by intramolecular dynamics of rhodanese in 5 M GdmCl.* (A) Short time correlation with linear time base including a fit to the data resulting in a time constant of 80 ns for intramolecular dynamics, in good agreement with the autocorrelation data (Fig. 2 main text). (Inset) Comparison of fits including (black) and excluding (green) distance dynamics. (B) Complete correlation function with logarithmic time base. Fits and inset analogous to (A).

## References

[1] F. Hillger, D. Nettels, S. Dorsch, B. Schuler, *J. Fluoresc.* **2007**, *17*, 759-765.

[2] A. L. Horwich, S. G. Burston, H. S. Rye, J. S. Weissman, W. A. Fenton, *Methods Enzymol.* **1998**, *290*, 141-6.

[3] A. M. Bhattacharyya, P. M. Horowitz, *J. Biol. Chem.* **2001**, *276*, 28739-43; J. A. Mendoza, E. Rogers, G. H. Lorimer, P. M. Horowitz, *J. Biol. Chem.* **1991**, *266*, 13044-13049; B. M. Gorovits, P. M. Horowitz, *Biochemistry* **1998**, *37*, 6132-5; G. Zardeneta, P. M. Horowitz, *J Biol Chem* **1992**, *267*, 5811-6.

[4] N. P. Huang, J. Vörös, S. M. De Paul, M. Textor, N. D. Spencer, *Langmuir* **2002**, *18*, 220-230.

[5] B. K. Müller, E. Zaychikov, C. Bräuchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508-3522.

[6] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 2655-2660.

[7] B. Schuler, *Methods. Mol. Biol.* **2007**, *350*, 115-38.

[8] G. Lipari, A. Szabo, *Biophys. J.* **1980**, *30*, 489-506; K. Kinosita, S. Kawato, A. Ikegami, *Biophys. J.* **1977**, *20*, 289-305; G. F. Schröder, U. Alexiev, H. Grubmüller, *Biophys J* **2005**, *89*, 3757-3770.

[9] J. B. Pedersen, *J. Chem. Phys.* **1972**, *57*, 2680-&.

[10] Z. S. Wang, D. E. Makarov, *J. Phys. Chem. B* **2003**, *107*, 5617-5622.

[11] B. W. Van Der Meer, Coker, G. III, Chen, S. Y. S., *Resonance energy transfer: theory and data*, VCH Publishers, Inc., New York, **1994**.

[12] Ü. Mets, in *Fluorescence Correlation Spectroscopy* (Eds.: E. S. Elson, R. Rigler), Springer-Verlag, Berlin, **2001**.

[13] P. Kask, P. Piksarv, M. Pooga, Ü. Mets, E. Lippmaa, *Biophys. J.* **1989**, *55*, 213-220.

# 4. FRET–trajectories of single immobilized Csps

## *Methods*

### Sample preparation

*Expression and purification of CspBc*

The plasmid pet11a, containing the coding sequence for Csp*Bc*, was obtained from Christine Magg and Franz Xaver Schmid (University of Bayreuth). Q2 and L66 were replaced by cysteine residues. The protein sequence is:
MCRGKVKWFNNEKGYGFIEVEGGSDVFVHFTAIQGEGFKTLEEGQEVSFEIVQGNRGPQ AANVVKC. The protein was expressed in *E. coli* BL21 (DE3) cells. Freshly transformed cells from a carbenicillin LB-plate were used to inoculate a 120 mL LB-preculture that was grown at 37 °C to an OD of 0.4. Six prewarmed 400 mL LB-cultures were inoculated with 20 mL of the precultures and were grown and shaken at 37 °C and 170 rpm. Expression was induced at an $OD_{600\ nm}$ of 0.8 with 1 mM IPTG, grown for two more hours and harvested. Cells were dissolved in 40 mM Tris, 2 mM EDTA, pH 7.0, and lysed by French press. Cell debris was removed by centrifugation for 45 minutes at 38000 g. To digest DNA and RNA, 5 µL (25 U/µL) benzonase (Novagen) were added per 25 mL supernatant and incubated for 30 minutes in the presence of 10 mM $MgCl_2$. After adjustment of the pH with sodium hydroxide to 6.2 to 7.0, bacterial protein was heat precipitated by heating the sample for 40 minutes to 65 °C. Precipitate was separated from supernatant by centrifugation at 38000 g for 30 minutes. The supernatant was adjusted to 2 mM EDTA and 5 mM mercaptoethanol and dialyzed over night against 50 mM sodium carbonate, 2 mM EDTA, 5 mM mercaptoethanol, pH 7.0. (Buffer was prepared with a pH of 7.0 but had a pH of 6.2 after sterile filtration due to removal of oxygen.) Anion exchange chromatography in the same buffer was used to remove remaining bacterial protein by loading the protein with a superloop on an equilibrated HiPrep 16/10 QFF column (GE Healthcare) and washing it with buffer. Csp*Bc* eluted during the washing step, whereas the bacterial protein bound to the column and was eluted with a gradient from 0 to 0.6 M sodium chloride. The eluate was checked by SDS-PAGE and fractions containing Csp*Bc* were pooled.

DNA and remaining protein were removed from Csp*Bc* by applying hydrophobic interaction chromatography. Pooled fractions were adjusted to the starting conditions of the chromatography by diluting them 1:2 in a doubly concentrated buffer and loaded on a butyl sepharose 4FF column (GE Healthcare) equilibrated with 38 % saturated ammonium sulfate, 50 mM sodium citrate, 5 mM EDTA, 5 mM mercaptoethanol, pH 6.2. The column was washed and Csp*Bc* was eluted with 50 mM sodium citrate, 5 mM EDTA, 5 mM mercaptoethanol, pH 6.2.

It was confirmed by UV absorption that DNA was removed. The sample was concentrated using an Amicon stirred ultra filtration cell (Millipore) with a membrane of a molecular weight cut off of 3500 Da. Gel filtration was done with a HiLoad 26/60 Superdex75 pg column (GE healthcare) in 50 mM sodium phosphate, 10 mM DTT, pH 7.0. Csp*Bc*-containing samples were combined and concentrated to 2 mg/mL or higher. The concentration was determined be UV absorption, TCEP adjusted to neutral pH was added to a concentration of 2 mM, and aliquots of 2 mg were stored at -20 °C.

### *Labeling of Csp*Bc

An aliquot of 2 mg Csp*Bc* was adjusted to 4 M GdmCl, 2 mM TCEP, incubated for one hour, and desalted with a HiTrap Desalting column (GE healthcare) in 50 mM sodium phosphate, pH 7.0. The samples were collected under argon atmosphere and the protein concentration of each fraction determined spectroscopically ($A_{280\ \mathrm{nm}\ (0.1\ \%)} = 0.97$). Fractions with a concentration of 1 mg/mL or higher were combined. The fluorophores Alexa Fluor 488 and 594 maleimide (Invitrogen) were freshly dissolved in DMSO (10 µg dye in 1 µL DMSO) and sonicated for 10 minutes. The donor (Alexa Fluor 488) was added in a 1:1 molar ratio to the protein and incubated at 4 °C over night. The singly labeled protein was separated from unlabeled and doubly labeled protein by chromatofocusing with a Mono P column (GE healthcare) using 25 mM BisTris buffer, pH 6.3, as start buffer and Polybuffer 74 as eluent. The column was equilibrated according to the manufacturer's instruction. The reacted protein was diluted 1:8 in start buffer and loaded on the column. The fractions containing singly labeled proteins, as confirmed by matrix-assisted laser desorption/ionization mass spectroscopy, were concentrated and adjusted to 50 mM sodium phosphate, pH 7.0. The singly labeled protein was reacted with the acceptor dye Alexa Fluor 594 in a 1.5-fold molar excess at 4 °C over night. Doubly labeled protein was separated from singly labeled protein and free dye by anion exchange chromatography with a Mono Q column (GE healthcare) using 10 mM sodium phosphate, 0.01 % Tween20, pH 5.0, as a start buffer and the same buffer with 1 M sodium chloride as eluent. A gradient was used from 0 to 60 % eluent, and the doubly labeled protein eluted at 40 %. The mass was again confirmed by mass spectroscopy. Labeled Csp*Tm* and proline peptides were obtained from Ben Schuler and were prepared as described previously (3).

### *Vesicle preparation*

In order to observe the folding and unfolding of individual Csp*Bc* molecules proteins were encapsulated in lipid vesicles. The vesicle preparation was done at the GdmCl concentration final needed, ensuring the same GdmCl concentration inside and outside the vesicles. Proteins were used at a concentration that yielded a suitable fraction of vesicles including one and a negligible fraction of vesicles with two or more protein molecules. Most of the vesicles were empty. The distribution of the empty, singly and multiply filled vesicles was calculated with a Poisson distribution, assuming a vesicle size of 100 nm.

Empty vesicles were prepared by extrusion of biotinylated lipids (egg-phosphatidylcholine and biotinylated phosphoethanolamine in a ratio of 500:1; Avanti lipids) in 50 mM sodium phosphate buffer at the appropriate GdmCl (Pierce) concentration and 0,0005 % Tween 20 (Pierce), pH 7.0, through a polycarbonate membrane with a pore size of 100 nm (Avestin, Mannheim, Germany) using a LiposoFast-basic extruder (Avestin, Mannheim, Germany). Protein-containing vesicles were prepared by extrusion of a mixture of 0.9 μM labeled protein, 5.5 mg/mL lipids, 0.001 % Tween 20 (Pierce), and 50 mM sodium phosphate, pH 7.0, with the appropriate GdmCl concentration.

### Ensemble experiments

#### *Protein stability*

For the unfolding transition of the unlabeled protein, stock solutions of Csp*Bc* with a concentration of 1 mg/mL were diluted into different GdmCl concentrations and 50 mM sodium phosphate, pH 7.0, to a final protein concentration of 10 μg/mL and incubated for 30 minutes at room temperature. Tryptophan fluorescence was measured with a Fluorolog 4. Emission Spectra were recorded from 290 to 450 nm, exciting the sample at 280 nm, using slit width corresponding to 5 nm for excitation and emission, an integration time of 0.5 s, and a wavelength increment of 2 nm. All spectra were corrected for the emission profile of the xenon lamp and the detection efficiency of the photomultiplier at different wavelengths. Spectra were corrected for the background from impurities of GdmCl used by assuming a linear dependency of the background on GdmCl concentration. A difference spectrum of the native sample and the sample measured at the highest GdmCl concentration was calculated. For the transitions, the signal at the wavelength with the maximum amplitude (356 nm) in the difference spectrum was used. Data were analyzed with a two-state model according to Santoro and Bolen (58).

Analogous to the tryptophan transition, a FRET transition of doubly labeled Csp*Bc* was measured. The protein stock solution with a concentration of 10 μg/mL was diluted to a final concentration of 0.1 μg/mL. Emission spectra were recorded, exciting at 490 nm, from 510 to 800 nm with excitation and emission slit widths corresponding to 5 nm, an integration time of 0.5 s, and a wavelength increment of 2 nm. The wavelengths with the maximum total amplitudes of the difference spectra were 522 nm and 612 nm for donor and acceptor, respectively.

To calculate FRET, the amplitudes were converted to the integrals below the donor and acceptor peaks with correction factors 48.24 (donor) and 78.25 (acceptor) that were determined from free dye spectra. Transition data were further analyzed as described for the tryptophan data.

*Stopped-flow experiments*

Kinetic measurements were done with a π*-180 stopped-flow spectrometer (Applied Photophysics). Labeled protein with a concentration of 1 µM in either 0 or 5 M GdmCl, 0.01 % Tween 20 and 50 mM sodium phosphate, pH 7.0, was diluted 11-fold with different concentrations of GdmCl, 50 mM sodium phosphate, pH 7.0. The sample was excited at 435 nm with a slit width of 4 nm, and fluorescence emission was recorded at wavelengths above 495 nm. Due to the weak detection efficiency of the photomultiplier tube at wavelengths larger than 600 nm, in the long wavelength range, mainly donor photons were recorded. The resulting transients were fit with single exponential decays. Folding rates obtained were plotted logarithmically against the GdmCl concentration (chevron plot) and fit with the equation

$$\ln k_{obs} = \ln\left(k_f \cdot e^{(-m_{kf}[D])} + k_u \cdot e^{(mk_u[D])}\right),$$

where $k_{obs}$ is the observed relaxation rate dependent on the denaturant concentration, $k_f$ is the folding rate, and $k_u$ is the unfolding rate under native conditions and therefore the intercept with the y-axis. $m_{kf}$ and $mk_u$ describe the sensitivities of the folding and unfolding rates, respectively, on the denaturant concentration $[D]$. With this equation, the transition midpoint and the folding and unfolding rates at the transition midpoint were determined (59).

**Single molecule experiments**

*Single molecule free diffusion experiments*

Observations of single molecule fluorescence were made using a MicroTime 200 confocal microscope (PicoQuant) equipped with a continuous-wave 488-nm diode laser (Sapphire 488–100 CDRH; Coherent) and an Olympus UplanApo 60/1.20W objective. Sample fluorescence was separated into donor and acceptor components using a dichroic mirror (585DCXR; Chroma Technology) and two final filters (ET525/50M and HQ650/100; Chroma Technology). See Fig. 1 for an overview. Each component was focused onto an avalanche photodiode (SPCM-AQR-15; PerkinElmer Optoelectronics), and the arrival time of every detected photon was recorded. Samples of labeled protein were diluted to a concentration of 20 pM in 50 mM sodium phosphate buffer at the appropriate GdmCl (Pierce) concentration, individually adjusted to pH 7.0. Tween 20 (0.001 %; Pierce) was added to prevent surface adhesion of the protein. The measurements were performed at a laser power of 100 µW. Successive photons detected in either channel separated by less than 100 µs were combined into one burst. Identified bursts were corrected for background, differences in quantum yields of the donor and acceptor, different collection efficiencies in the detection channels, cross-talk, and direct acceptor excitation, as described by Schuler (60). A burst was retained as a significant event if the total number of counts exceeded 50.

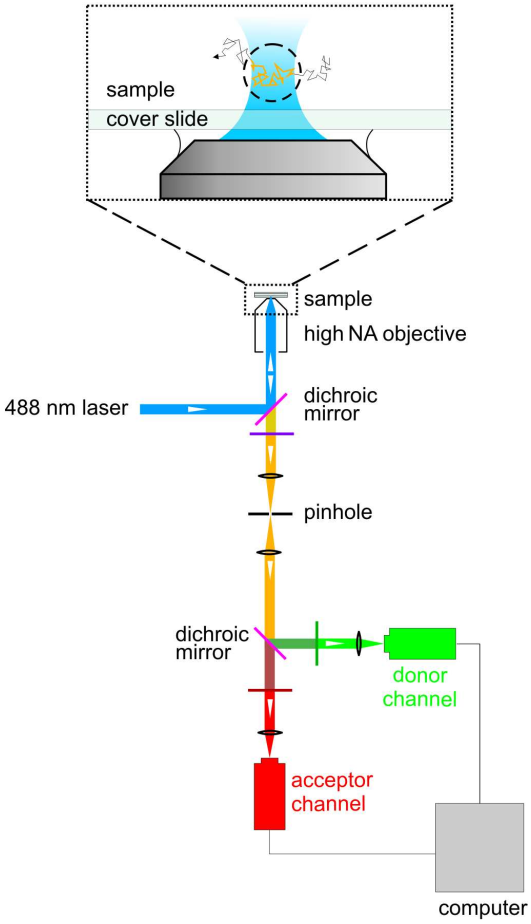

Fig. 1. Schematic representation of the instrumentation for confocal single molecule measurements adapted from Schuler (8). White triangles in the light path indicate the propagation direction of light. Blue is excitation light, red and green are fluorescence light and yellow the overlay of both. The dotted box shows an enlargement of objective and sample. See the text for details.

### *Preparation of the flow cell*

The quartz flow cell with a volume of 50 µL was rinsed with 1 mL buffer. A lipid bilayer was applied to the quartz flow cell by fusion of the vesicles upon rinsing with 50 µL of the plain (biotinylated) vesicle solution, incubated for 5 minutes, and rinsed with buffer. The Csp-vesicles were coupled to this surface by avidin-biotin linkage. For this purpose, 50 µL of a 1 mg/mL solution of avidin was applied, incubated for 5 minutes, and rinsed. Protein-loaded vesicles were applied after a 1:100 dilution, incubated for 5 minutes, and rinsed with buffer. The same GdmCl concentration was used in all solutions. The fluorescence background was recorded after each step after rinsing the sample. The flow cell in the sample holder is shown in Fig. 2.

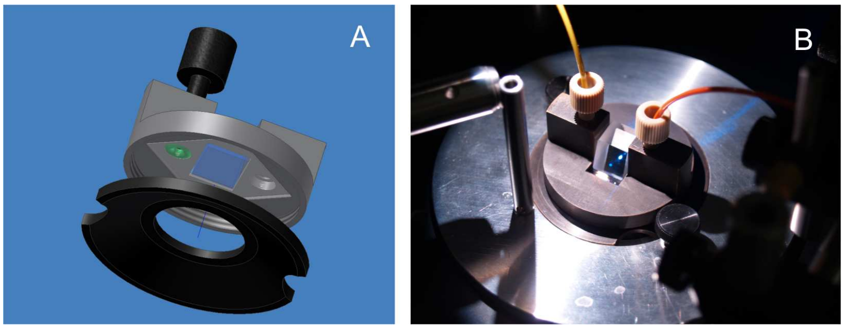

Fig. 2. Flow cell and sample holder with TIRF microscopy. (A) Technical drawing of the sample holder designed by Dominik Hänni and Daniel Nettels (University of Zurich). Top black part is a chromatography connector for tubing, green is a rubber o-ring that seals the connection between tubing and glass with hole. The light gray square is the cavity for the flow cell. (B) Photo of the sample holder with flow cell, TIRF prism, and inlets.

### *Confocal measurements of immobilized proteins*

The sample was observed by confocal measurements with the same instrumentation as described for single molecule free diffusion experiments. The sample was moved with a piezo stage to scan an area of 15 µm × 15 µm at 100 µW. Each identified molecule, visible as a bright spot, was then accessed individually with the piezo stage and measured at a power of 10 µW until bleaching of the dyes. Trajectories were recorded with the software SymphoTime (PicoQuant) and analyzed with Mathematica (Wolfram Research).

Donor-only labeled or vesicle-encapsulated Csp*Bc* was immobilized on the surface, and polarization trajectories were recorded. A similar instrumentation as for confocal FRET measurements was used. Additionally, the excitation light was circularly polarized by using a λ-quarter wave plate, and instead of a dichroic mirror, a polarizing beam splitter cube was used to split the fluorescence light into components horizontally ($I_h$) and vertically ($I_v$) polarized with respect to the excitation light. The fluorescence polarization ($P$) was calculated from $I_h$ and $I_v$ according to

$$P = \frac{I_v - GI_h}{I_v + GI_h}$$

with the factor $G$ correcting for the different sensitivities of the detectors. The polarization was calculated for every time frame (100 ms) of each molecule. From the trajectories, only fragments with a total intensity corresponding to the intensity of a single dye (60 to 130 photons per 100 ms) were used. The polarization of the time frames within these fragments was averaged for each trajectory.

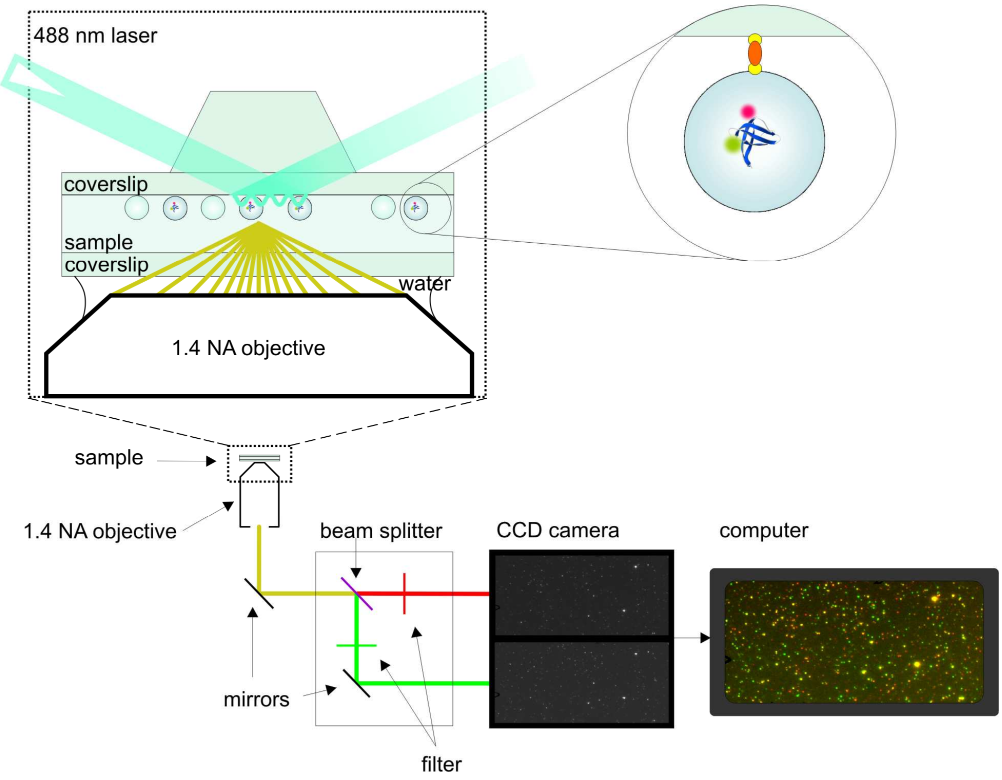

Fig. 3. Schematic representation of the instrumentation for TIRF microscopy measurements. The exciting light is depicted in blue, fluorescence light in yellow, green, and red. The magnification shows a vesicle with protein bound to the surface of the cell. See the text for details.

*TIRF measurements of immobilized proteins*

The sample was observed by prism-type TIRF microscopy with an inverted microscope. The objective used was an Olympus UplanApo 60x/1.20W. The sample was excited with a continuous-wave 488-nm diode laser (Sapphire 488–100 CDRH; Coherent) at the absorption maximum of the donor with 35 mW, measured before the prism. The fluorescence image of an area of 100 µm x 50 µm was separated by a beam splitter and filters (ET525/50M and HQ650/100; Chroma Technology) into a red and a green channel and imaged on the chip of a back-illuminated EMCCD-camera with an image of 512 x 512 pixel (Ixon-EMCCD, Andor, Belfast, N. Ireland). See Fig. 3 for a schematic overview. Movies with a length of 40 s and a frame rate of 10 Hz were recorded in the "fits"-format. Up to 300 molecules could be measured per movie.

## Data analysis

The automated identification of immobilized molecules from the recorded movies was done using the software TIRFikon developed by Daniel Nettels (University of Zurich). All 400 images taken per movie were averaged per pixel. From this time-averaged image, the background was subtracted and molecules were selected if their brightest pixel had an intensity value between 1000 and 3000. The corresponding fluorescence trajectories of the selected molecules were extracted from the movie. Because the emission from each molecule was distributed over several pixels, their intensity was averaged over all selected pixels per frame. The further analysis was done using Mathematica. $E$ for each trajectory was calculated, and the trajectories were searched for jumps.

*Jump identification and trajectory fragment selection*

Trajectories were excluded from the analysis if the fluorescence intensity of one dye exceeded 3000 or the sum of both dyes exceeded 3500 in at least one time frame (100 ms). The trajectories were searched for anti-correlated change points of donor and acceptor signal (Fig. 4A). A moving average was calculated separately for acceptor and donor traces over seven frames (700 ms)[1] to minimize scatter. From the binned data obtained (Fig. 4B), identification trajectories (Fig. 4C) that report about the position of change points by peaks were calculated. Differences between fluorescence intensities of neighboring frames with different step sizes were added as described here in detail.

To identify the change points in donor or acceptor signal, an identification trajectory $c(i)$ is calculated for each signal, where $i = 1 \dots n$ denotes the $i$th time frame and $n$ is the total number of frames. The total number of frames is $n$. For each frame $i$, the value $c(i)$ of the identification trajectory is calculated according to (see Fig. 4D for illustration):

$$
\begin{aligned}
&c(1) = 0, c(2) = 0,\\
&c(i) = |f(i) - f(i+1)| + |f(i+1) - f(i-1)| + |f(i-1) - f(i+2)|\\
&\qquad\qquad + |f(i+2) - f(i-2)|,\\
&c(n-1) = 0, \text{ and}\\
&c(n) = 0.
\end{aligned}
$$

Here, $f(i)$ is the intensity of the $i$th time frame. The resulting identification trajectory was used to locate the jumps in the fluorescence trajectory. If the calculated value $c(i)$ was higher than a threshold of 1000 or the sum of four succeeding values was higher than 600, the signal change in donor or acceptor, respectively, was identified as a change point. The criterion of the *anti-correlated* change point was fulfilled if the donor and simultaneously the acceptor signal changed

---

[1] The averaging of 700 ms was only used to identify the change points; further analysis was done with non-binned data.

with opposite sign and a minimum of 200 in the fluorescence intensity. For Csps, only trajectory fragments next to an anti-correlated change point were selected, and the resulting data were used for the histograms.

For CspBc, the jump identification method was additionally used as described without binning in order not to miss any events of the faster folding protein. In that case, not only the surrounding four neighboring frames, but the surrounding eight were used for the jump identification in order to be able to find transitions additionally to the short timescale of below 400 ms in the range of 400 to 800 ms.

In addition to the search for anti-correlated change points, $\langle E \rangle$ and the standard deviation in $E$ of every trajectory fragment between identified change points (not necessarily anti-correlated) were calculated. Additionally, $\langle E \rangle$ and standard deviation were calculated for the fragment between the beginning and the first change point as well as between the last change point and the end of the trajectory. Bleached parts of the trajectories could be largely excluded by using only fragments with a standard deviation of $E$ smaller than 0.15. The standard deviation in donor and acceptor fluorescence intensity within a fragment had to be below 250. By this condition, slow intensity changes within one fragment could be largely excluded. Fragments were selected if the sum of donor and acceptor fluorescence intensity was in the range of 800 and 2500.

Finally, the intersection of the set of fragments identified by standard deviation and the set of fragments identified by anti-correlated change points was used. Different values for the identification via brightness in TIRFikon, thresholds for selection criteria used in Mathematica, such as for anti-correlated jumps and standard deviation were tested for the different proteins and their suitability judged from the separation of peaks in $E$ histograms and from the separation of change point populations in transition maps.

### *Trajectories of donor-only labeled CspTm*

Trajectories of donor-only labeled proteins were analyzed with criteria different from those of FRET-labeled proteins previously described. Donor-only trajectories with fluorescence intensities higher than 1500 were discarded. The mean of the fluorescence intensity of every fragment had to be below 1500. The maximum allowed standard deviation of $E$ was not limited. The standard deviation of the donor fluorescence intensity had to be below 250 to select the fragment for analysis. By setting the above-mentioned conditions, the signal of bleached dyes and dark states was included in the analysis and could be characterized. The frequency of jumps to or from dark states could be observed, whereas in the FRET analysis only non-bleached trajectory fragments are taken into account.

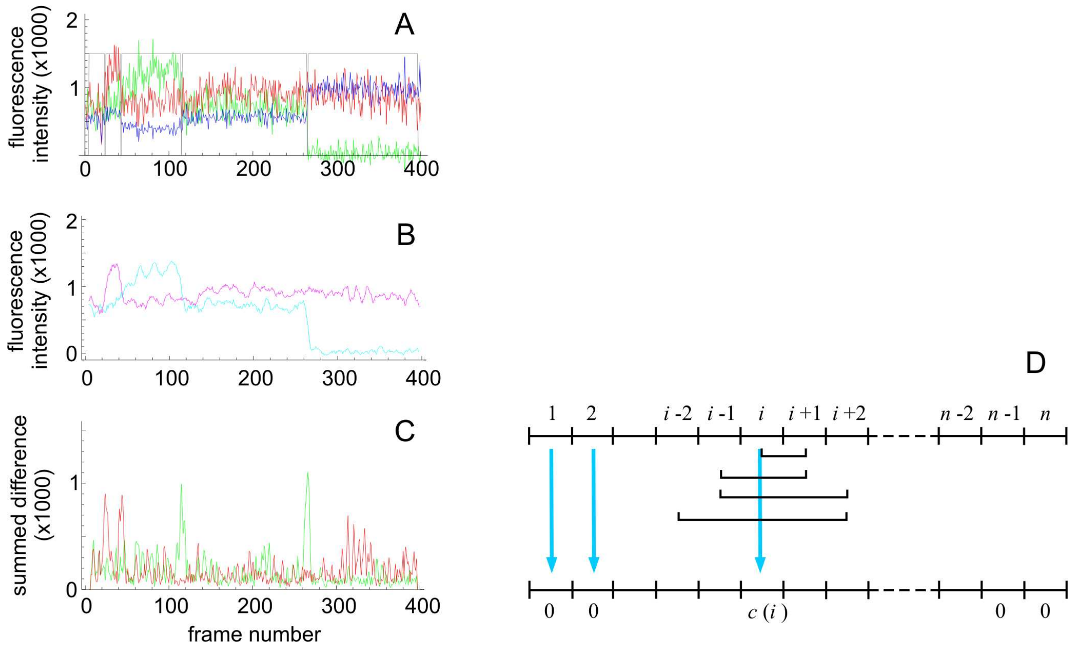

Fig. 4. Change point identification in fluorescence trajectories. (A) Donor fluorescence (green), acceptor fluorescence (red), and E (blue) recorded with a bin size of 100 ms. The gray line indicates that fragments are selected, where the value is 1500 and shows change points where the value is 0. Change points are additionally indicated by purple points. (B) Binned signal of donor (cyan) and acceptor (purple) fluorescence trajectory from A. (C) Identification trajectory of donor (green) and acceptor signal (red) of the fluorescence trajectory in A. Purple points indicate the position of a change point. (D) Schematic of the change point identification procedure. The upper row shows the time frames 1 … $n$ of the binned trajectory. The lower row displays the identification trajectory. Each value is calculated from the binned trajectory (blue arrows). The ends of the four horizontal brackets indicate the frames from which the difference in fluorescence intensity was calculated. The sum of the four differences $c(i)$ yields the value of time frame $i$ for the identification trajectory. Details are described in the text.

# Results

## Characterization of labeled Csp*Bc*

Csp*Bc* was labeled with donor and acceptor dyes at the termini of the peptide chain. In order to confirm two-state behavior of the protein with introduced cysteines and to determine the conformational stability of the labeled protein, the folding midpoint was measured by kinetic and equilibrium experiments.

For the denaturant-dependent equilibrium transition, FRET between donor and acceptor was probed, and a folding midpoint of 2.3 M GdmCl was obtained (Fig. 5A). For the unlabeled protein, containing cysteine residues, but no dyes, and using tryptophan fluorescence intensity as a probe, a transition midpoint of 1.8 M GdmCl was obtained (Fig. 5A), which corresponds to a

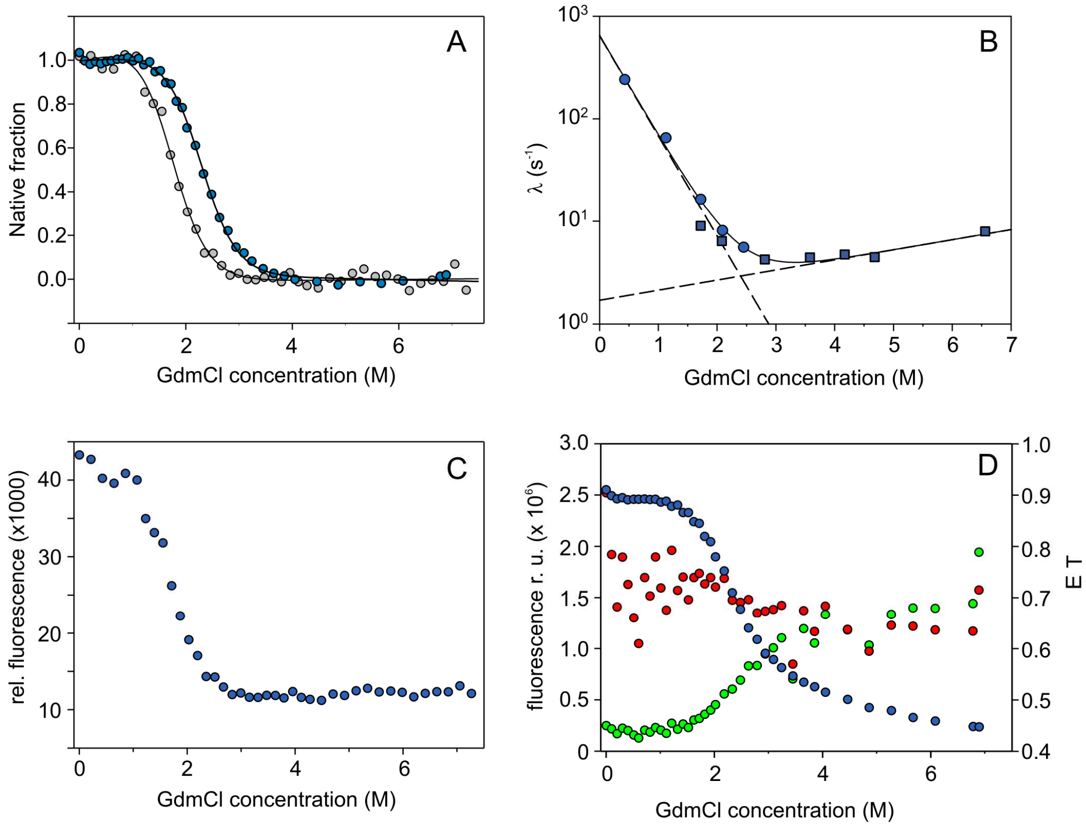

Fig. 5. Denaturant dependent equilibrium transition of Csp*Bc*: (A) Transition based on tryptophan fluorescence of the unlabeled protein (gray) and $E$ of the labeled protein (blue). Fluorescence signals are normalized to fraction native. Transition midpoints were at 1.8 M and 2.3 M for tryptophan and FRET-based transition, respectively. Continuous lines represent least-squares fit analysis to the experimental data based on a 2-state folding mechanism. (B) Chevron plot of Csp*Bc*: Folding (circles) and unfolding rates (squares) of the FRET labeled protein dependent on the final GdmCl concentration. The transition midpoint was at 2.4 M GdmCl, the folding rate at the transition midpoint was 2.9 s$^{-1}$. The fit results for folding and unfolding rates (dashed line) and the sum of the rates (solid line) are shown. (C) Raw data of the transition monitored by tryptophan fluorescence. (D) Raw data of the transition monitored by dye fluorescence and FRET. Shown are donor fluorescence (green), acceptor fluorescence (red), and FRET (blue).

destabilization of 4.2 kJ/mol. Raw fluorescence data of tryptophan and FRET measured transition are shown in Fig. 5C and Fig. 5D, respectively.

For the determination of folding kinetics, stopped flow experiments were used. Again, FRET between the fluorophores was measured, in this case by monitoring the donor intensity (see methods). The Chevron plot shows a transition midpoint of 2.4 M GdmCl (Fig. 5B), in good agreement with the equilibrium experiments. The folding and unfolding rate at the folding midpoint is 2.9 $s^{-1}$.

## Stochastic folding simulations of single molecules

The folding rate of 2.9 $s^{-1}$ should allow the detection of several transitions during the observation time in immobilization experiments. The observation time is 40 s in TIRF experiments but is limited in many cases by bleaching of the dyes. In previous experiments with Csp*Tm,* only few transitions could be observed until the dyes bleached within this time. Csp*Tm* folds and unfolds with a folding rate of 0.4 $s^{-1}$ (3). The different folding frequencies of the two Csps are visible in stochastic simulations (Fig. 6). With the higher folding rate of Csp*Bc*, one order of magnitude higher than that of Csp*Tm*, ten times more transitions should be observed before bleaching of the dyes occurs.

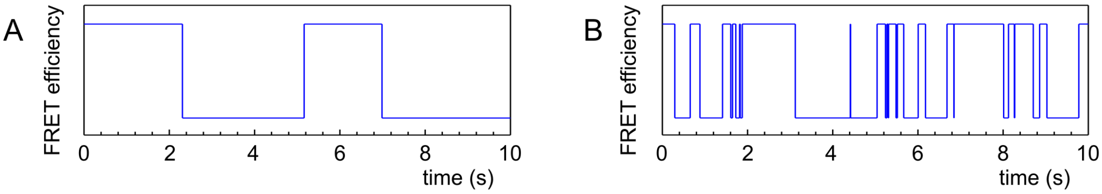

Fig. 6. Stochastic simulations of the folding frequency of Csp*Tm* and Csp*Bc* at the unfolding midpoint, where Csp*Tm* folds and unfolds with a rate of 0.4 $s^{-1}$ (A) and Csp*Bc* folds and unfolds with a rate of 2.9 $s^{-1}$ (B). A two-state system with a low FRET efficiency in the unfolded and high FRET efficiency in the folded state was assumed.

## Single molecule diffusion experiments

Csp*Bc* was observed in free diffusion single molecule experiments in order to compare FRET efficiencies to those measured under immobilizing conditions. The obtained $E$ histograms of free diffusion single molecule experiments show two clearly separated populations with the native population at $E \approx 0.95$ and the denatured population with $E$ depending on the GdmCl concentration. With increasing denaturant concentration, the native population decreases while the denatured population increases. For Csp*Bc*, $E$ of the denatured state shifts from 0.75 at low GdmCl concentrations (0.5 M and 1 M) to 0.4 at 6 M GdmCl (Fig. 8A), representing an expansion, as seen in Fig. 8B. At 2.5 M GdmCl concentration, the native state is populated to a similar extent as the unfolded state based on peak areas.

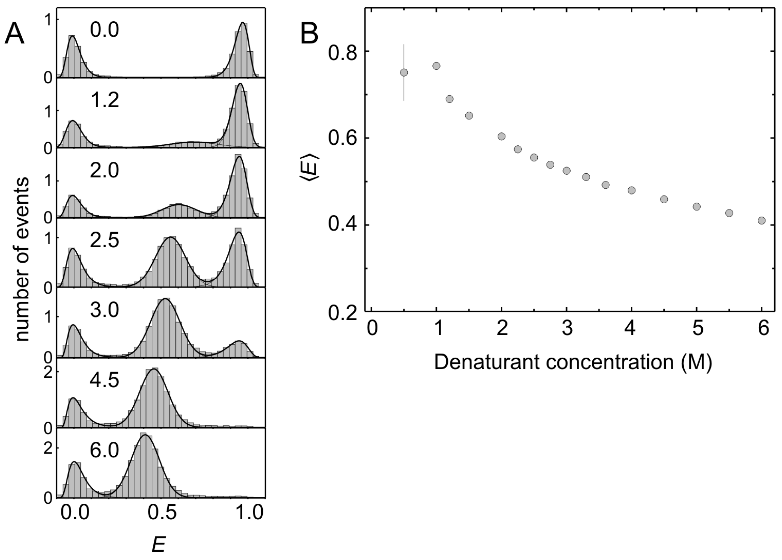

Fig. 8. Single molecule diffusion experiments: (A) $E$ histograms of CspBc. The GdmCl concentration in M is indicated in the panels. (B) $\langle E \rangle$ of the unfolded population of Csp*Bc* dependent on GdmCl concentration obtained from $E$ histograms. The error bar indicates the standard deviation estimated from two measurements. For GdmCl concentrations larger than 1.2 M, the error in $E$ is estimated to be ~0.02. Only for low GdmCl concentrations, the unfolded population cannot be fitted accurately because of its small population.

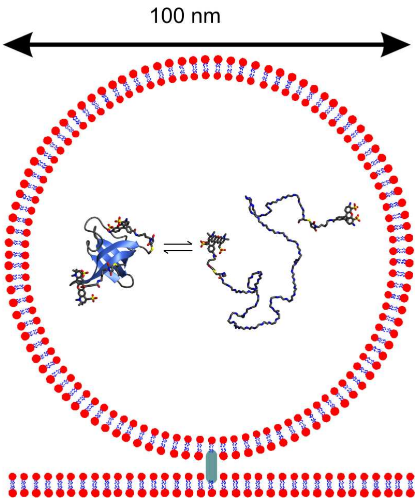

Fig. 7. Schematic representation of a protein encapsulated in a vesicle immobilized on a surface. The average diameter of the vesicles is 100 nm. The vesicles are bound to the quartz surface that is coated with a lipid bilayer via avidin-biotin chemistry. Relative dimensions are not to scale. Picture is taken from Rhoades *et al.* (44).

## Single molecule immobilization experiments

For observing individual protein molecules for extended periods of time, proteins were encapsulated into vesicles, and the vesicles were attached to the surface to keep them in the confocal volume, or in the same position for TIRF measurements, respectively. Since the protein is much smaller than the vesicle, it is expected that it can diffuse freely inside the compartment and can fold as in free solution. A schematic drawing is shown in Fig. 7.

Initial experiments were done using confocal detection. Immobilized molecules were visible as bright spots on a scanned area of the sample. Each molecule had to be accessed individually and measured until the fluorophores bleached, which was highly time consuming. To improve statistics, a TIRF-setup was used (Fig. 3), where many molecules could be observed at the same time. However, trajectories with anti-correlated donor and acceptor signal—as expected for a folding or unfolding transition—could be observed only very rarely (Fig. 9). The majority of the fluorescence trajectories showed changes in the donor and acceptor signal without obvious correlation. There were several different types of behavior observed: many of the anti-correlated

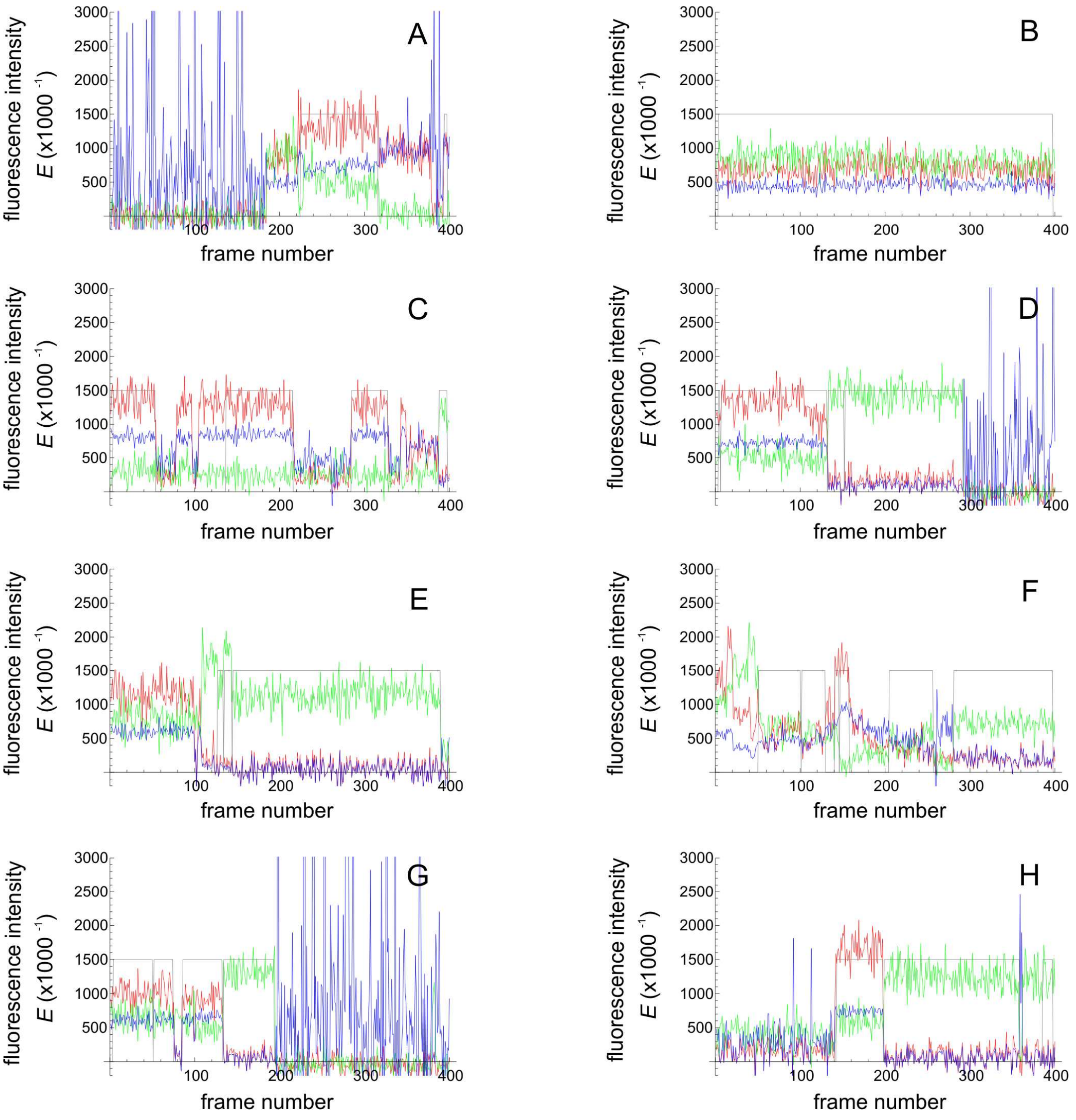

Fig. 9. Trajectories of Csp*Tm* at 2 M GdmCl. Trajectories in this figure are selected to show different events observed. Red and green traces are fluorescence intensities of donor (green) and acceptor (red), blue is the calculated $E$. Purple dots indicate the identified jumps (occurring in donor or acceptor signal or in both). The gray trace indicates if the part of the trajectory was used for the analysis (value of 1500) or not (value of zero).

transitions could be assigned to acceptor bleaching. In these cases, acceptor fluorescence is decreasing with a simultaneous increase of donor fluorescence because transfer can no longer take place. In some trajectories, following acceptor bleaching, also donor bleaching is visible within the 40 seconds of recording time. In trajectories that showed step-like transitions for one of the dyes, the second dye showed in many cases behavior not correlated to the first dye. Dark states of one of the dyes that did not affect the other dye were observed and lasted up to a few seconds. Additionally, slow transitions without a step-like jump from high to low fluorescence signals were observed for donor and acceptor signals. Some trajectories showed a decrease of the fluorescence signal in multiple steps, indicating that more than one protein was enclosed in a vesicle. The large number of trajectories that did not show simple step-like and anti-correlated signal suggests that other events such as dye photophysics take place besides folding and unfolding. A selection procedure was developed to extract the events related to folding.

Although the trajectories do not exhibit simple anti-correlated behavior, the overall color pattern of Csp*Bc* at different GdmCl concentrations corresponds qualitatively to the expected population of folded and unfolded conformations. This behavior can be seen by comparing the time-averaged TIRF images of 40 s-measurements of Csp*Bc* at different GdmCl concentrations in Fig. 10: in low GdmCl concentrations (0.1 M), many molecules show red (acceptor) fluorescence, indicating high $E$ as expected for native proteins. The averaged signal changes with intermediate (2.0 M) GdmCl concentrations to yellow (indicating both green and red fluorescence) and with high concentrations (3.5 M) to green (donor) fluorescence corresponding to lower $E$, i.e. unfolded proteins.

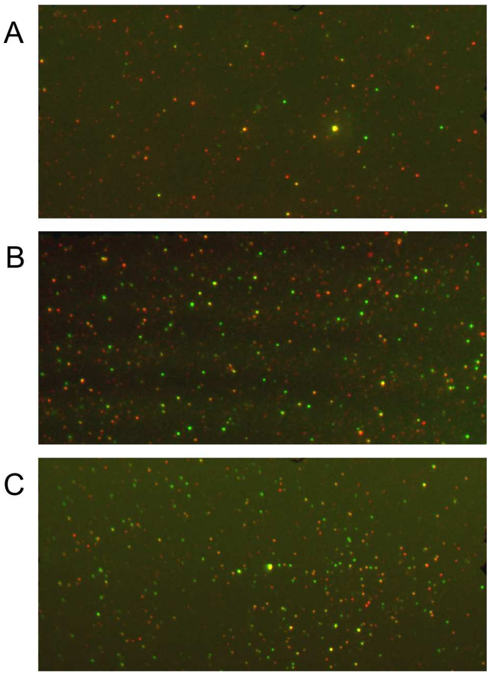

Fig. 10. Time-averaged images over 40 s TIRF measurements of immobilized Csp*Bc*. The panels shows an area of 100 µm x 50 µm of the flow cell surface. Acceptor emission is shown in red, donor emission in green. Yellow is the overlay of both. Panels show samples at 0.1 M (A), 2.0 M (B), and 3.5 M (C) GdmCl concentrations.

*Signal selection process*

In order to determine whether folding processes contribute to the signal changes rather than fluorophore photophysics, an automated analysis was performed that avoids a manual selection process that can yield strongly biased results. The analysis is described in more detail in the Methods section.

Fluorescence trajectories were selected with the software TIRFikon and further processed with Mathematica. $E$ was calculated, and different criteria had to be met by the trajectories or their fragments to be selected for the analysis. First, the trajectories were searched for anti-correlated jumps in donor and acceptor signal, since this is an indication for a folding or unfolding transition. Additionally, trajectory fragments between the jumps were analyzed according to further criteria. For every fragment, the mean and standard deviation were calculated for $E$ and donor and acceptor intensity over the whole fragment. The standard deviation of $E$ proved to be a good criterion to identify bleaching because the apparent $E$ fluctuates strongly if both dyes are bleached and their intensity is close to zero. Fragments were excluded from the analysis if acceptor and donor were bleached at the same time because fragments contain scatter that would contribute to a large width of $E$. Fragments were only selected for further analysis if the standard deviations were below a certain value to exclude fragments showing slow (not jump-like) intensity changes, because slow changes are not expected in trajectories presenting folding. Trajectories showing signal with high total fluorescence intensity, indicating more than one protein in the vesicle, and trajectories without anti-correlated signal were discarded, too. Since the protein encapsulation is a stochastic process, vesicles filled with multiple proteins are expected. From the Poisson distribution of proteins in vesicles under these condition, it is estimated that in the experiments presented here, 20 % of the filled vesicles are filled with two proteins (61).

From the trajectories, only fragments were selected that fulfilled both the standard deviation criterion and the anti-correlated jump criterion of donor and acceptor signal. From these selected parts of the trajectories, $E$ histograms and transition maps were plotted.

$E$ histograms were plotted to compare the results of the immobilization experiments and free diffusion experiments and to find defined folded and unfolded populations. $E$ of every time frame (100 ms) of the selected signal fragments measured at the same GdmCl concentration were used for the histogram, i.e., the longer a fragment is, the more does it contribute to the histogram. As a result, there is a significant shot noise contribution in the $E$ histograms.

In transition maps , all pre and post-jump values of $E$ of the trajectories were plotted with $E$ before the jump on the abscissa and $E$ after the jump on the ordinate, respectively. In contrast to $E$ histograms, transition maps report on $\langle E \rangle$ of the fragments before and after a jump equally, independent of the fragment length.

### *Donor-only labeled Csp*

Due to the high complexity observed in the trajectories, also donor-only-labeled trajectories of CspTm were recorded to check whether the donor dye alone shows complexity higher than expected. Among the selection criteria to include the fragment into the analysis was the fluorescence brightness, which had to be within the intensity limits for one molecule (corresponding to two dyes). Jumps were recognized by the jump analysis, which is in this case only based on the change of the donor intensity since an acceptor is not present. The trajectory fragments obtained were further selected using the standard deviation criterion. Since selection for anti-correlated signal in donor and acceptor fluorescence could not be used, the donor intensity changes in Fig. 11 are only based on the standard deviation criterion.

Fluorescence trajectories show different types of intensity changes. Many trajectories were observed where the dyes exhibited blinking. A quantitative analysis of the donor-only labeled Csp*Tm* revealed that many fluorescence intensity jumps are present in these experiments, even though no change of fluorescence intensity is expected other than bleaching.

Bleaching contributes to the lower right half of the transition map, where most jumps are located. A large jump population changes in fluorescence intensity from 1200 to values around zero and a smaller population changes from fluorescence intensity of 1200 to 700. Additionally, the opposite changes from low to high fluorescence intensities are present (from values of 100 to 1100 and from values of 600 to 1200) that represent the return from dark states. 493 jumps were identified after standard deviation criteria were applied and were divided by the number of selected trajectories (670). The resulting average number of jumps per trajectory is 0.7. It is expected that additional complexity is added to the system by using additionally an acceptor dye.

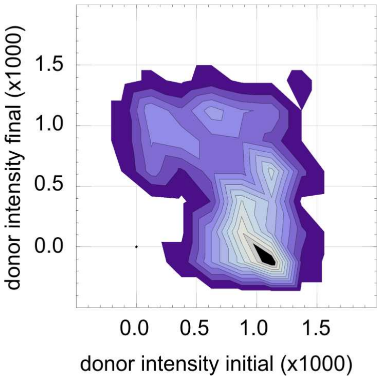

Fig. 11. Transition map of the donor fluorescence intensity of donor-only labeled Csp*Tm*. All changes in fluorescence intensity that were recognized by the automated analysis were plotted with their values before (abscissa) and after a jump (ordinate) are included on the plot.

### *Different transfer efficiencies in immobilization and diffusion experiments*

There is a discrepancy in $E$ observed between diffusion and immobilization experiments for most proteins observed. Pro$_{20}$ (Fig. 12) and the unfolded population of Csp*Tm* (Fig. 13) show larger values of $E$ in immobilization experiments than in diffusion experiments (3, 40, 62). In contrast to that, the unfolded population of Csp$Bc$ at 0.1 M GdmCl shows smaller $E$ in immobilization experiments than in diffusion experiments. On the other hand, $E$ of $\mathrm{Pro}_6$ and of the native population of Csp$Tm$ is in good agreement for both kinds of experiments.

Thus, the discrepancy in $E$ cannot be solely due to the different instrumentation used in TIRF experiments in comparison to the confocal instrumentation for the diffusion experiments because the variation has opposite sign dependent on the sample. The origin of this discrepancy could not yet be clarified. However, for the kinetic studies of proteins, the calibration to exact values of $E$ is not crucial and therefore $E$ was not corrected in TIRF experiments.

*Polyproline peptides*

Proline peptides were measured to test the suitability of the system. The peptides used contained either 6 prolines ($\mathrm{Pro}_6$) or 20 prolines ($\mathrm{Pro}_{20}$) and were labeled at the cysteine introduced at the C-terminus and at the N-terminal glycine residue. For proline peptides, trajectories without anti-correlated donor and acceptor fluorescence intensity during the jumps were included in the analysis to quantify the heterogeneity caused by dye photophysics photochemistry. Bleaching of the dyes can occur in the trajectories, and hence should be visible in the transition maps. In all $E$ histograms of the immobilization experiments, the donor-only peak is visible with $\langle E \rangle$ of 0.1 (Fig. 12). This peak results from molecules whose acceptor dye is bleached or temporarily in a dark state. The donor-only peak in diffusion experiments is always close to zero. The reason why it is higher in TIRF immobilization experiments must be due to the acceptor fluorescence that is not zero in the donor-only fragments. This could happen if the acceptor background was not completely subtracted, but is unlikely in this case, because it was taken care to subtract the backgrounds of donor and acceptor fluorescence to the same level. It is also possible that both backgrounds are not completely subtracted. In that case also higher values of $E$ are expected. Shotnoise and a heterogeneous background additionally broaden the peaks but do not shift the position.

In single molecule free diffusion experiments, $\mathrm{Pro}_6$ showed a peak at $E \approx 0.95$ (62). The same value is observed in $E$ histograms of immobilization experiments on $\mathrm{Pro}_6$ presented here. In transition maps, a change of $E$ from 0.95 to 0.2 is seen, which corresponds to acceptor bleaching (Fig. 12C). In addition to the high $E$ and the donor-only population, a small population with $E \approx 0.55$ is seen in $E$ histograms and transition maps, which is populated from $E \approx 0.95$. A small population is seen that corresponds to bleaching with $E$ changing from 0.55 to values below 0.2.

If only trajectories of $\mathrm{Pro}_6$ with anti-correlated jumps are selected (Fig. 12B), as it is done for the analysis of Csps, the population at $E \approx 0.55$ increases in the $E$ histograms relatively to the population $E \approx 0.95$. Although this selection is not suitable for proline peptides, since no folding or other change in conformation is expected, it shows that also in proline peptides some events are occurring that give a transition-like signal. These events cannot be related to folding because it is known from free diffusion single molecule experiments that only one population besides the donor-only population is present. The origin of this additional population cannot be explained clearly but might be due to dye photophysics.

$\text{Pro}_{20}$ showed a peak with $E \approx 0.7$ in $E$ histograms (Fig. 12A and B) and transition maps (Fig. 12C), which converts to $E$ corresponding to the acceptor-bleached population. The observed $E \approx 0.7$ is different from the value of 0.55 observed in single molecule free diffusion experiments (62) and shows a rather broad distribution in comparison to $\text{Pro}_6$. The separation between the

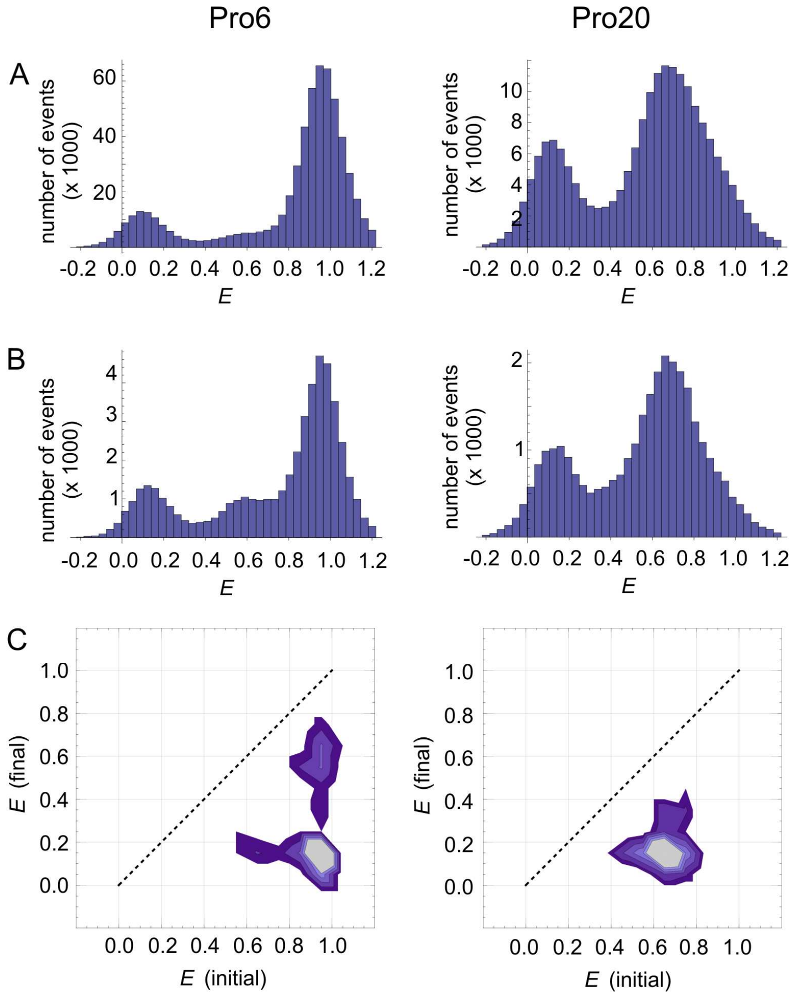

Fig. 12. $E$ histograms and transition maps from single molecule immobilization experiments of $\text{Pro}_6$ (left) and $\text{Pro}_{20}$ (right). $E$ from every time frame selected from the trajectories was used for the histograms. (A) Trajectories with anti-correlated and non-anti-correlated signal were used. (B) Only trajectories with anti-correlated signal were used. (C) Transition maps of proline peptides. Changes in $E$ before (abscissa) and after a jump (ordinate) are plotted. All jumps are included in the plot that met both criteria, the anti-correlated jump criterion and the standard deviation criterion. The number of jumps increases from dark blue to light blue; black areas are highly populated areas that are off-scale. Peaks were cut to allow a finer resolution for the lower frequency occurrences.

donor-only and the $Pro_{20}$-peak is slightly better if the non- anti-correlated signal is included in the analysis (Fig. 12A). This is again an indication that the trajectories contribute to anti-correlated signals with $E$ different from the main population with $E \approx 0.7$ and different from acceptor-bleached molecules.

Proline peptides show mainly bleaching from their characteristic $E$, which is visible as populations in the lower right half of the transition maps. For both proline peptides, bleaching ($E$ changes to values below 0.2) is visible. For $Pro_6$, transitions to possibly dark states (towards $E \approx 0.55$) are seen additionally.

### *Cold Shock Proteins (Csps)*

For Csps, only trajectories showing an anti-correlated fluorescence signal were chosen from the immobilization experiments because anti-correlated jumps are expected if folding takes place. As for the proline peptides, the donor-only peak is present for the Csps at all GdmCl concentrations. The histograms (Fig. 13 and Fig. 14) of the selected signal from the trajectories of Csp*Bc* and Csp*Tm* show different distributions of populations dependent on the GdmCl concentration. These differences in the distributions are also seen in the transition maps, but interestingly, they do not show distinct populations that correspond to folding or unfolding events for both Csps. As a comparison, $E$ from every time frame individually of all measured trajectories without discarding any signal is shown in Fig. 13C. For all proteins, the majority of the changes of $E$ is located in the lower right half of the transition map, which indicates that $E$ is lowered and proteins are unfolding or acceptor dyes are bleaching.

### *CspTm*

Although fluorescence trajectories of immobilized proteins are very heterogeneous, the $E$ histograms of immobilization experiments are reflecting the differently populated states depending on the GdmCl concentration approximately as known from free diffusion experiments. The three $E$ histograms obtained from immobilized molecules of Csp*Tm* (Fig. 13B) show approximately a bimodal distribution, but it is not as clearly defined as for the single molecule diffusion experiments (3, 40). The $E$ histograms of immobilization experiments differ from the histograms obtained by single molecule diffusion experiments in some points.

In diffusion experiments, mean measured $E$ for the unfolded state is smaller in comparison to the data of immobilization experiments. For free diffusion experiments, mean measured $E$ of the unfolded state is 0.45 at 1 M GdmCl (40), whereas it is about 0.7 for immobilized molecules (Fig. 13B). The native population is located at $E$ of 0.95 for both diffusion and immobilization experiments. The unfolded population is very small in diffusion experiments, whereas it is quite prominent in immobilization experiments. At 2 M GdmCl in $E$ histograms of immobilization experiments, one main population with $E \approx 0.65$ is visible, and additionally a weak shoulder with $E \approx 0.9$, indicating the native population, whereas in diffusion experiments, $E \approx 0.43$ for the unfolded population and $E \approx 0.95$ for the native state. In diffusion experiments, the native fraction is still observed, whereas it is barely visible in immobilization experiments. Only at a concentration of 1 M GdmCl, the native state is populated significantly. As it is already seen in the $E$ histograms of immobilization experiments, transition maps show hardly any native population at 2 M GdmCl. This may be an indication towards a destabilization under the immobilization conditions, since in free diffusion experiments there is still a prominent native population visible at 2 M GdmCl. In diffusion experiments in 3 M GdmCl, measured $E$ is 0.4, whereas it is 0.6 in immobilization experiments.

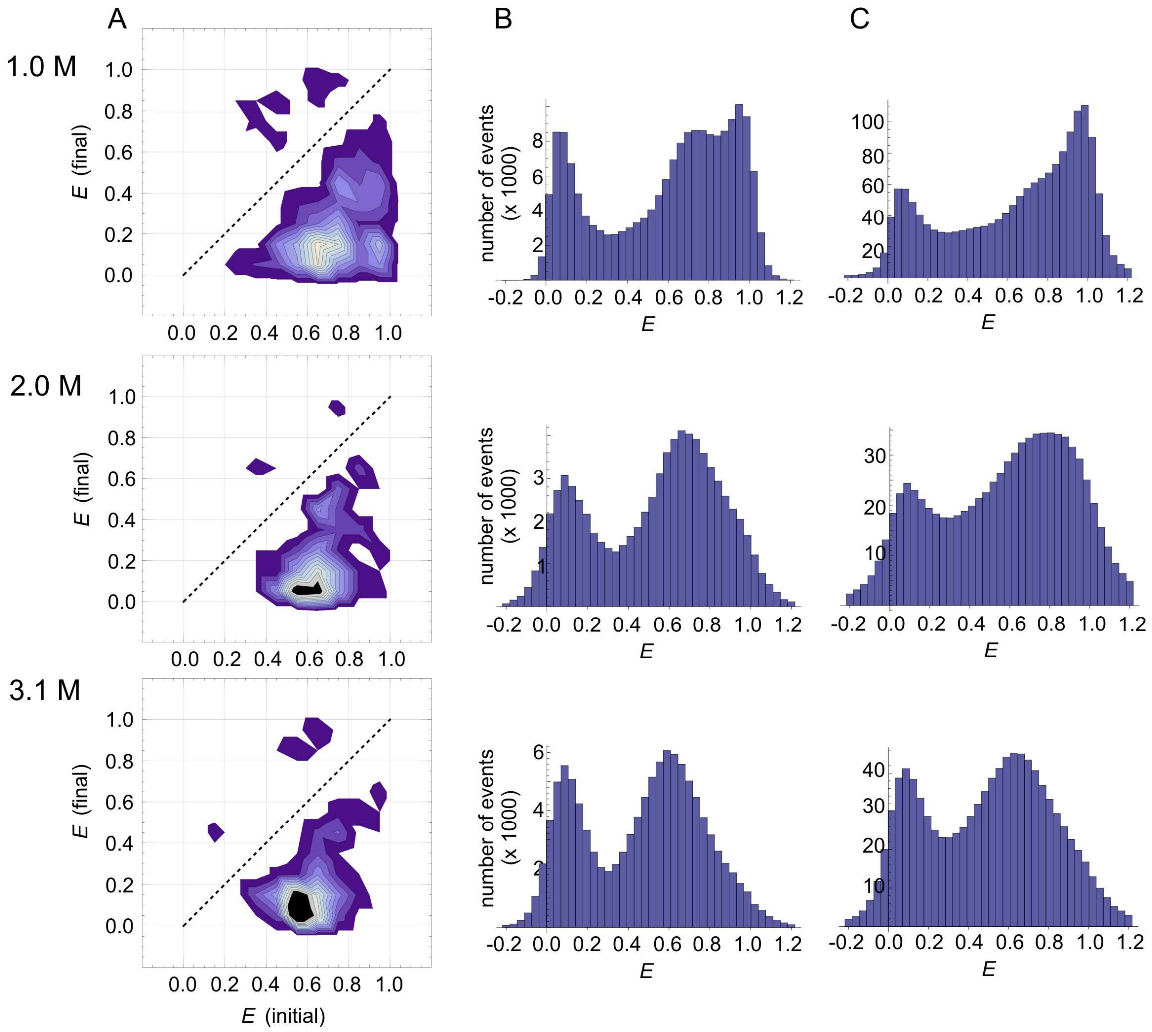

Fig. 13. Transition maps and $E$ histograms of Csp*Tm* in 1.0, 2.0, and 3.1 M GdmCl as indicated. (A) Transition maps of all jumps meeting both, the anti-correlated jump criterion and the standard deviation criterion are shown. (B) Only trajectories with anticorrelated signal were used for the $E$ histograms. (C) $E$ histograms of all trajectories including all fragments measured.

The transition maps indicate at all three GdmCl concentrations a change from the unfolded state with $E$ between 0.55 and 0.7, dependent on the GdmCl concentration, to the donor-only population with $E < 0.2$, which corresponds to acceptor bleaching (Fig. 13A). The advantage of the transition maps is that subpopulations of jumps can be resolved in some cases. At 1 M GdmCl, in addition to the jump population that changes from the unfolded to the donor-only population there is one minor population with a change from $E \approx 0.95$ to $E < 0.2$ that represents acceptor bleaching from the native state. Furthermore, two minor populations change from the native state with $E \approx 0.95$ and from $E \approx 0.75$ to $E \approx 0.45$. The change of $E$ from 0.95 to 0.45 may correspond to a transition from the folded to the unfolded state with the same $E$ as in diffusion experiments at 1 M GdmCl. Therefore, it seems that most of the molecules in immobilization experiments behave differently from free diffusing molecules. The origin of the transition from $E \approx 0.75$ to $E \approx 0.45$ is unclear. At 2 M GdmCl, no clear subpopulations are found and the unfolded population is rather one broad distribution that covers all observed $E$ values from 0.4 to 0.8.

At all three GdmCl concentrations measured, there are few events located in the upper left half of the transition map. Some of these events, those with $E$ changes from the unfolded state (0.55 to 0.7) to 0.95 (native state), could be assigned to folding events and are present at all GdmCl concentrations, but these represent the minority of all events. The corresponding unfolding could be seen clearly only for changes from $E \approx 0.95$ to values between 0.55 and 0.7 at 1 M GdmCl.

*CspBc*

$E$ histograms of the immobilization experiments of Csp*Bc* show one peak rather than both native and unfolded populations (Fig. 14B) as observed in free diffusion experiments of Csp*Bc* (Fig. 8).

At 0.1 M GdmCl, a main population with transfer $E \approx 0.95$, corresponding to the native population, is visible in $E$ histograms (Fig. 14B) and transition maps (Fig. 14A). The separation between the donor-only and the native peak is very low, but the transition maps show a defined population at $E \approx 0.55$. In diffusion experiments, the unfolded population is only visible at 1.2 M or higher GdmCl concentrations and is located at $E \approx 0.65$ at 1.2 M GdmCl. It is expected to be located at even higher $E$ at lower GdmCl concentrations. Interestingly, in immobilization experiments, the unfolded conformations are already populated at almost native conditions at 0.1 M GdmCl (Fig. 14B). Transition maps indicate that acceptor bleaching is occurring from the native and the unfolded state. Additionally, one population changing from $E \approx 0.95$ to $E \approx 0.55$ is present, possibly corresponding to unfolding events.

At 2.3 M GdmCl, there is one broad peak with a $E \approx 0.65$ in the $E$ histogram of immobilized molecules and in transition maps, corresponding to the unfolded population, whereas

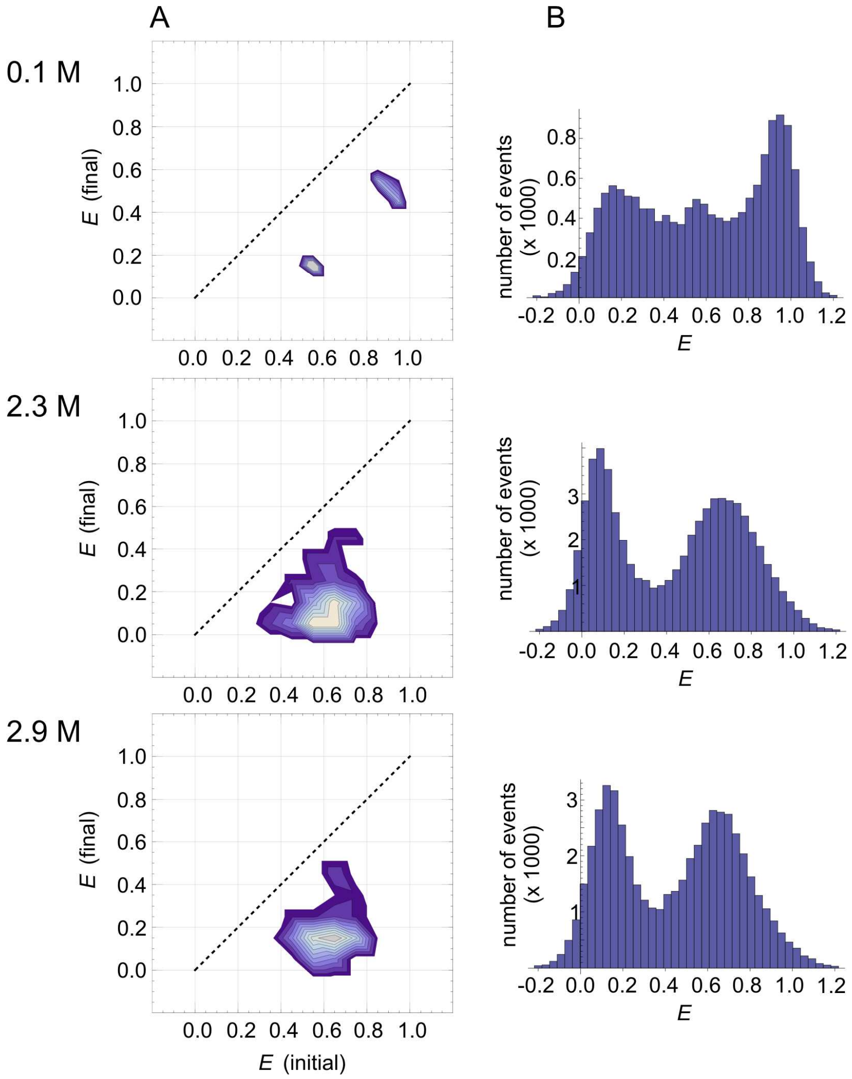

Fig. 14 . Transition maps and $E$ histograms of Csp*Bc* at 0.1, 2.3, and 2.9 M GdmCl as indicated. (A) Transition maps of all jumps meeting both, the anti-correlated jump criterion and the standard deviation criterion are shown. (B) Only trajectories with anticorrelated signal were used for the $E$ histograms.

in diffusion experiments the unfolded peak is at $E \approx 0.55$ at the same GdmCl concentration (Fig. 8). In diffusion experiments, a large native population relative to the unfolded population is visible, whereas it is almost not present in immobilization experiments. This is a further indication for destabilization in immobilization experiments in comparison to diffusion experiments, as also suspected from the Csp*Tm* measurements. Similar as for Csp*Tm*, in the transition maps only bleaching and unfolding are observed.

At 2.9 M GdmCl, there is a peak with $E \approx 0.65$ in $E$ histograms and transition maps for immobilization experiments, compared to $E \approx 0.55$ of the unfolded peak in free diffusion experiments. As for 2.3 M GdmCl, no folding transitions, but only bleaching is visible (Fig. 14A).

The transition maps, which allow an easier identification of subpopulations than in $E$ histograms, show that almost no native protein is present in immobilization experiments at 2.3

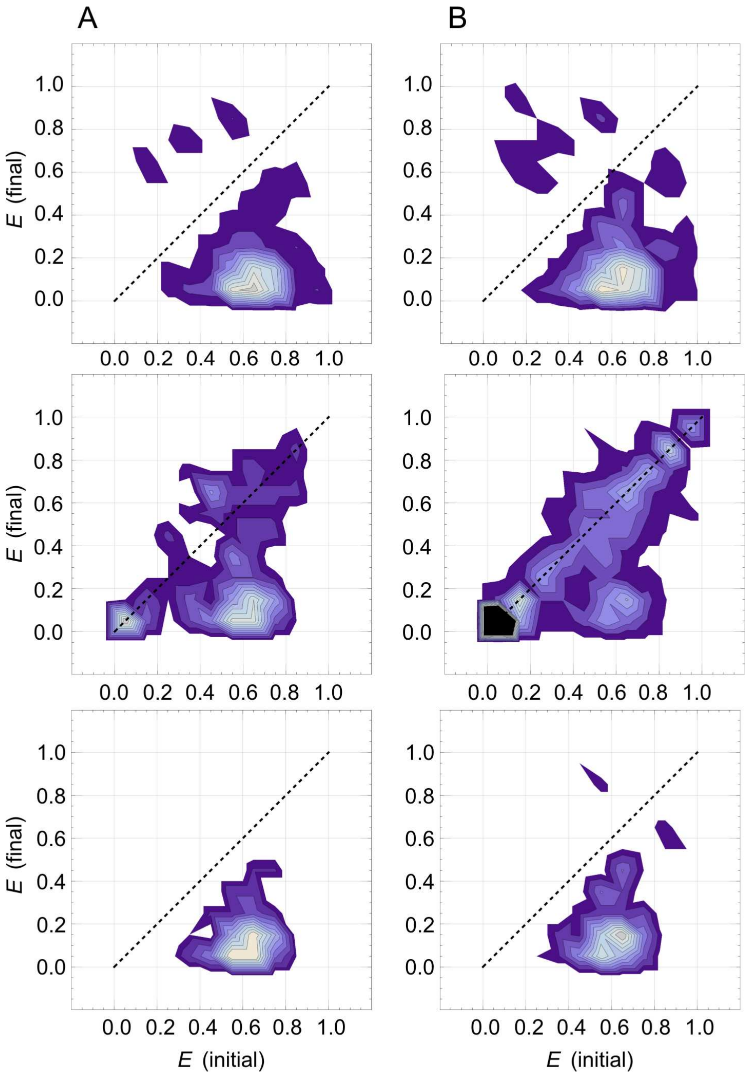

Fig. 15. Transition maps of Csp*Bc* at 2 M GdmCl depending on different binning of the data used for the jump identification. (A, left) Binning of 7. (B, right) Binning of 1. Upper panels show the signal meeting anti-correlation criterion, middle panels show jumps meeting the standard deviation criterion, and lower panels show the intersection of jumps meeting both criteria.

and 2.9 M GdmCl. Therefore, the presence of only one peak for Csp*Bc* in the $E$ histograms is not only due to a bad separation between native and unfolded peak, but reflects the populations present at the chosen GdmCl concentrations.

$E$ of the unfolded population is shifting from $E \approx 0.55$ at 0.1 M GdmCl, to $E \approx 0.65$ at 2.3 M GdmCl and 2.9 M GdmCl. This is the opposite direction as expected for the equilibrium collapse, but the shift is so small that it is within the uncertainty of the measurement.

As a reason for the missing peak separation, the time resolution of the experiments of 100 ms used with the CCD camera was considered. With a folding rate of 2.9 $s^{-1}$ for Csp*Bc*, the average folding time of Csp*Bc* is 345 ms. For the identification of the jumps, a binning of 700 ms was used. Hence, it is possible that some folding events are not identified, if they occur in rapid succession. Regarding the few folding events found overall (also for Csp*Tm*), it is rather unlikely that two events are happening at the same time.

However, the jump identification was repeated without binning for control according to the description in the methods section for CspBc. The transition maps (Fig. 15) derived from these identifications did not differ significantly. It is seen that many jumps are found meeting the standard deviation criterion, but are then discarded if the intersection of both criteria is calculated and the resulting jumps found are very similar to the jumps found after the 700 ms binning was applied. Therefore, the reason of the missing separation of populations cannot be due to the faster folding rate of Csp*Bc* in comparison to Csp*Tm*.

*Polarization Experiments*

As shown by Boukobza *et al.* (45), polarization experiments of singly immobilized molecules provide information about the rotational freedom of the proteins in lipid vesicles. Donor-only labeled Csp*Bc* molecules were excited with circularly polarized light and the horizontally and vertically polarized fluorescence light emitted from each molecule was measured. For each molecule, the polarization was calculated per time frame and averaged per trajectory. The distribution of the averaged polarization values is shown in (Fig. 16).

If a molecule is sticking to the lipid vesicle surface, it is not able to rotate freely and is expected to show a fixed non-zero polarization. The signal of many molecules hindered in rotational freedom would therefore result in a broad distribution of polarization values. For freely rotating molecules, it is expected that the polarization is averaged during the sampling time and a narrow polarization distribution is expected. The narrow distribution of the averaged polarization with a full width at half maximum of 0.15 indicates that the molecules can freely rotate (45).

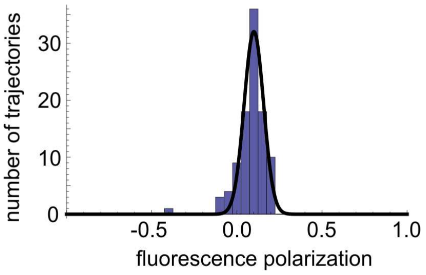

Fig. 16. Distribution of fluorescence polarization of individual donor-only labeled Csp*Bc* molecules immobilized in lipid vesicles. The solid curve is a Gaussian fit to the experimental data.

## Overview of the selection process and statistics

Fig. 17A shows the number of trajectories chosen for the analysis of the immobilization experiments, with different selection criteria. 18 to 37 % of the trajectories identified by TIRFikon were discarded due to high total fluorescence, indicating multiple proteins in a vesicle. If in addition non-anti-correlated trajectories are excluded from the analysis, there were finally 14 to 28 % of the initial trajectories (identified by TIRFikon) left for the Csps and 8 to 14 % for proline peptides (Fig. 17B), respectively. Of the trajectories finally chosen (yellow), only those fragments were selected for the analysis that agreed with the further criteria of anti-correlated jumps and standard deviation Fig. 18A.

The number of recognized jumps in all measured trajectories is given Fig. 18A. The number of jumps decreases significantly if selection criteria are applied to trajectories containing at least one anti-correlated jump. 34 to 74 % of the jumps in the selected trajectories meet the anti-correlation condition and 38 to 75 % meet the standard deviation criterion. In the case of CspBc under conditions without binning, very high numbers of jumps are found (75 to 78 %). The intersection of both criteria results in a number of selected jumps of 17 to 34 % of the identified jumps in the selected trajectories. It is seen also in transition maps that the jumps following anti-correlation and standard deviation criteria overlap only partially and therefore the resulting number of jumps is significantly smaller (Fig. 15). This relatively small overlap is mainly due to the identification of jumps that can be based either on a change of the donor fluorescence or the acceptor fluorescence, but not necessarily on both of them. Therefore, in the transition maps with the jumps selected according to the standard deviation criterion, there is a large jump population included that shows only a very small or no change in $E$ and is located on the diagonal. That is especially the case for CspBc with the jump identification without binning.

The order of applying the different selection criteria does not have an influence on the result. If trajectories without anti-correlated jumps are initially included in the analysis, a higher number of jumps pass the standard deviation criterion. During application of the anti-correlated jump criterion, these jumps are discarded and the identical jumps as with the method described above are resulting.

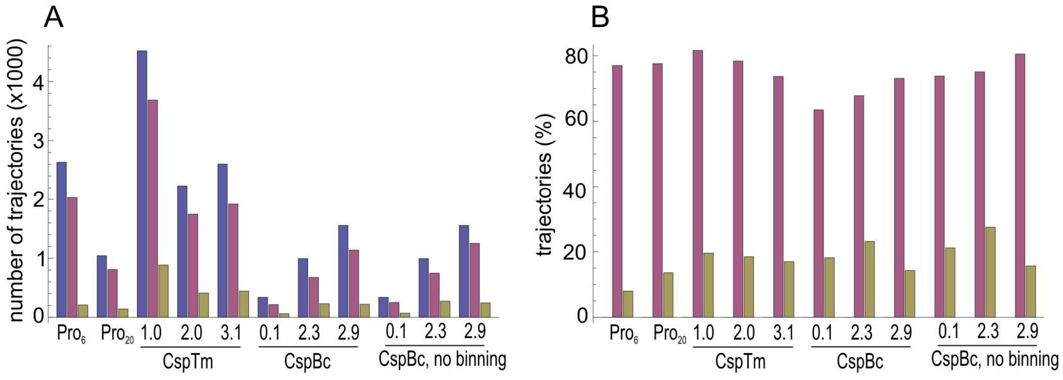

Fig. 17. Evaluation of trajectory statistics: (A) Number of trajectories selected after applying different selection criteria. All trajectories identified by TIRFikon (blue), trajectories left after signal with multiple molecules in one vesicle was excluded (purple), and after additionally non-anti-correlated trajectories were discarded and only trajectories containing at least one anti-correlated jump were kept (yellow). (B) Same as in A, but in percent of all identified trajectories.

## Frequency of jumps

Trajectories were analyzed for their jump frequency. The average number of jumps per trajectory was calculated for the different steps of the selection process. Only trajectories were used for the calculation that were finally selected and contained trajectories showing anti-correlated signal (Fig. 19A) or trajectories containing both anti-correlated and non-anti-correlated signal (Fig. 19B). Compared are the following jumps per trajectory: all identified jumps, including jumps only occurring in donor or acceptor signal, anti-correlated jumps, and jumps contained in the intersection of anti-correlated and standard deviation recognized jumps. The latter are therefore anti-correlated jumps that are between two adjacent fragments that passed the standard deviation criterion.

If the jump frequency of the different proteins measured in immobilization experiments would correspond to their folding behavior known from single molecule diffusion experiments, only a few jumps are expected for proline peptides. Diffusion experiments showed only one population besides the donor-only peak. Therefore, only jumps due to acceptor bleaching are expected. Some more jumps are expected for Csp*Tm*, which folds at the transition midpoint with a rate of 0.4 $s^{-1}$ (3), and even more for Csp*Bc*, which folds with a rate of 2.9 $s^{-1}$. In contrast to the measured folding rates, the anti-correlated trajectories (Fig. 19A) show no clear difference in the jump frequency between the two Csps, no matter which selection of jumps is used.

Additionally, a change in the folding rate and therefore in the average number of jumps per trajectory is expected with the GdmCl concentration for the Csps. If two 2-state folders, like Csps, follow the kinetic rates revealed from ensemble experiments also in single molecule experiment s, the highest folding and unfolding rate should be found at the transition midpoint.

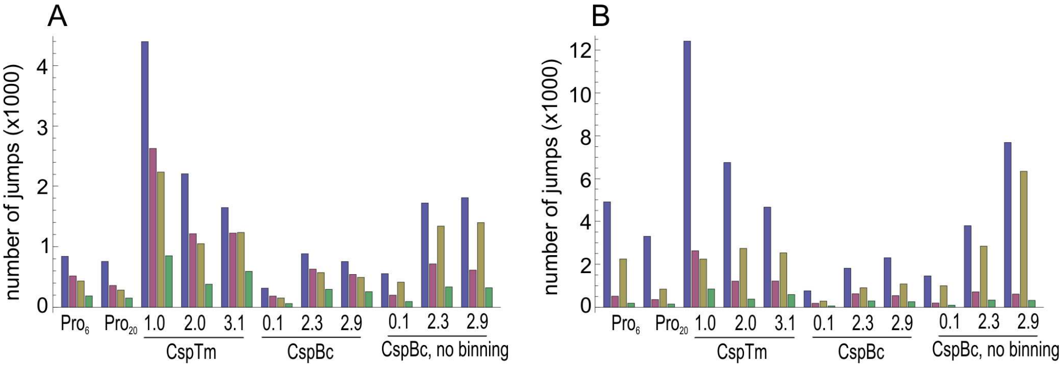

Fig. 18. Number of jumps in trajectories: (A) Number of jumps identified according to different criteria in trajectories showing at least one anti-correlated jump, (B) Jumps in non-anti-correlated and anti-correlated trajectories together. Jumps were selected according to the following criteria: all jumps identified, including jumps that show only a jump in donor or acceptor signal (blue), only anti-correlated jumps (purple), only jumps (not necessarily anti-correlated) that are next to fragments that fulfill the standard deviation criteria (yellow), and all jumps that are in the intersection of the latter two (green).

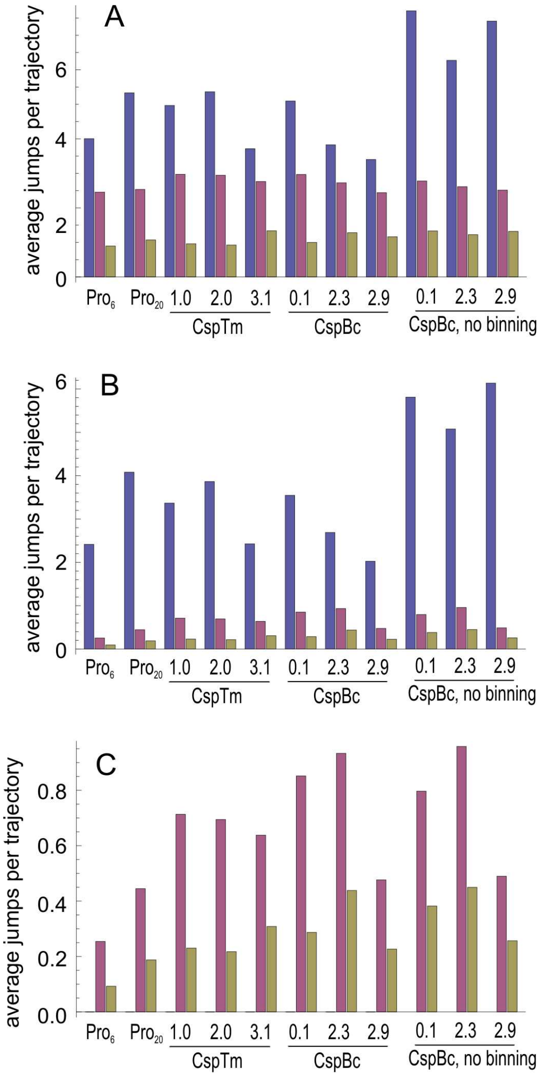

Fig. 19. Average number of jumps per trajectory. (A) Anti-correlated trajectories only, (B) anti-correlated and non-anti-correlated trajectories. (C) Enlargement of anti-correlated jumps and intersection from (B). All trajectories were used whose fragments and jumps were selected in the analysis. All jumps found initially (blue), jumps with anti-correlated signal (purple), and jumps selected with the intersection of anti-correlated and standard deviation criteria (yellow) were divided by the number of all selected trajectories.

According to the ensemble kinetic data, at this concentration the folding and unfolding rates are the same. At lower and higher denaturant concentrations, the folding rate of a single molecule should be lower because the unfolding rate under concentrations below the midpoint is slower than at the transition midpoint and hence limiting the folding frequency. The same effect of slower folding is expected for the folding at GdmCl concentrations higher than the transition midpoint. In this case, the folding rate is limiting. Especially at 3 M GdmCl for the Csp*Bc*, the folding rate is extremely low (Fig. 5B). In contrast to this consideration, differences in the frequency of jumps per trajectory between the different GdmCl concentrations of each Csp are not visible (Fig. 19).

For both Csps, the number of jumps per trajectory decreases from anti-correlated jumps to the intersection of anti-correlated jumps and jumps meeting the standard deviation criterion to a similar extent (Fig. 19A and C). Therefore, it is assumed that the standard deviation is not increased by partly captured folding events of Csp*Bc*, which would be the case if the identification method would not be able to find every jump, e.g. due to too strong binning of the data. In that case, jumps would be located within fragments, increase the standard deviation, and would be excluded from the analysis. In contrast to that, even more trajectories of Csp*Tm* are discarded because they do not meet the standard deviation criterion.

If only trajectories with anti-correlated jumps are considered (Fig. 19A), Csps and proline peptides are expected to show a difference in the number of jumps per trajectories, since prolines are assumed to show only acceptor bleaching events with anti-correlated signal. Csps are expected to display additionally to bleaching folding events, which should lead to a higher number of jumps per trajectory. However, no clear difference in the number of jumps per trajectory is seen for proline peptides in comparison to Csps, independent of the jump selection process (all jumps identified, anti-correlated jumps, or jumps according to the intersection criterion).

If the number of jumps per trajectory is compared, taking also trajectories without anti-correlated signal into account, there is not a clear difference between prolines and Csps considering all trajectories (Fig. 19B). If only anti-correlated jumps per trajectory are considered, instead of all jumps, the frequency for all proteins and selection criteria decreases by one order of magnitude. Although the jump frequencies of proline peptides are about half the jump frequency of Csps, they are still higher than expected and therefore not likely related solely to folding behavior. If only the intersection of anti-correlated jumps and jumps selected according to the standard deviation criterion is selected, the difference between proline peptides and Csps is less pronounced. $Pro_6$ shows 0.1, $Pro_{20}$ 0.2 jumps per trajectory, and the values for Csps are between 0.2 and 0.4 and therefore slightly higher.

This raises the question if the standard deviation criterion is helpful to select the folding related signal, and on the other hand how much folding related signal at all is contained in the trajectories. If from the anti-correlated jumps, only those are selected that meet additionally the standard deviation criterion, less than half of the jumps are remaining (Fig. 19B). Since proline peptides are not expected to show conformational transitions on the observed time scale, the anti-correlated events found in proline trajectories must be caused by dye photochemistry (or acceptor bleaching). Proline peptides show only slightly less anti-correlated jumps per trajectory than Csps. Since for Csps besides blinking and acceptor bleaching also folding events are expected, it is surprising that relatively few anti-correlated jumps are found for Csps in comparison to proline peptides.

From Fig. 17B it is seen that proline peptides display less trajectories with anti-correlated jumps than Csps. For $Pro_6$ and $Pro_{20}$, 8 % and 14 % of the trajectories, respectively, showed anti-correlated signal of all trajectories in the selected brightness range. Of Csps, 14 to 23 % contained anti-correlated jumps. The tendency of proline peptides displaying fewer transitions is also seen here, but again a stronger discrepancy was expected.

On average, 4 to 5 jumps per trajectory were found (Fig. 19A and B), but only about one was found per trajectory after intersection criteria were applied. If the transition maps are considered, it is seen that most of the anti-correlated signal is caused by bleaching of the dyes, but not visible in ranges of the plot where folding or unfolding would be expected. The frequency analysis is an additional hint that the trajectories and the transitions do not represent solely folding, but probably photochemical events of the dyes.

## Discussion

Previous work on vesicle encapsulated Csp*Tm* generated results with short trajectories and relatively low statistics (44). In that work, the bleaching of the fluorophores was limiting the recording time of the fluorophores to several seconds. Additionally, it was quite time consuming to measure every molecule individually.

In order to observe the folding of individual molecules more quantitatively, in this work Csp*Bc* was used, which folds according to a 2-state model, but folds and unfolds with a tenfold higher rate than Csp*Tm*. In the observation time, more transitions should be recorded for a better statistical analysis. Furthermore, proteins with different folding rates and stiff polyproline peptides were chosen. TIRF microscopy was used, which allowed the measurement of up to 300 molecules at a time. The large heterogeneity observed in the trajectories could be reduced by an automated analysis selecting only trajectories or fragments of trajectories containing constant signal between anti-correlated jumps (for Csps). However, the results obtained from the immobilization experiments are ambiguous. On the one hand, agreement was found in the equilibrium behavior averaged over time, but on the other hand, no significant differences between the Csp*Bc* and Csp*Tm* could be found in single molecule experiments in terms of the frequency of folding, in contrast to what is known from ensemble experiments.

### Qualitative agreement in histograms

$E$ histograms obtained from a large number of trajectories agree qualitatively with the histograms obtained from single molecule experiments with freely diffusing protein. As in diffusion experiments, $E$ histograms show two distinct populations for both Csps: the unfolded and the folded population. Although the measured $E$ of the unfolded state and the distribution of the peak size is not agree, the tendency that the native population is decreasing and the unfolded population is increasing with increasing GdmCl concentrations is also present in immobilization experiments. Immobilization experiments of proline peptides that are not folding due to their rather rigid structure were therefore used as a control and were found to agree qualitatively with free diffusion experiments. $\text{Pro}_6$ displayed the same $E$ as in free diffusion single molecule experiments, and only showed a small additional population whose origin is unclear. $\text{Pro}_{20}$ showed different $E$ in $E$ histograms of immobilized molecules and in free diffusion single molecule experiments. The origin of this discrepancy could also not be identified in the present work.

The partial agreement between immobilization and free diffusion experiments suggests that folding information is contained in the immobilization trajectories, but in addition, other events contribute to the signal.

### Unexpectedly few folding events and jumps with anti-correlated fluorescence signal

Assuming that the Csp trajectories show a signal related to folding and unfolding in addition to acceptor bleaching, they should contain a significantly higher number of anti-correlated jumps per trajectory than proline peptides, whereas Csp*Bc* trajectories are expected to contain more jumps than Csp*Tm* trajectories, as known from ensemble experiments. The examination of the frequency of anti-correlated jumps per trajectory revealed that the obtained kinetics do not correspond to those measured in ensemble experiments. Furthermore, they do not show a significant difference between proline peptides and Csps. There is no difference visible between the two Csps, although they fold with tenfold different folding rate according to ensemble experiments. Because for none of the proteins, trajectories with anti-correlated jumps were obtained that corresponded to the folding rates known from ensemble experiments, it seems that the measured signal is not related to folding, but almost independent of the measured sample. The concern that the folding events of Csp*Bc* cannot be resolved due to the high frequency of folding and too large binning times for the jump identification was excluded. The results suggest that the observed signal is rather a result of the dye properties, i.e., their transitions to dark states, but not signal related to folding.

Generally, the relatively high number of jumps without anti-correlated signal indicates that only one of the dyes is changing the signal, which would be not expected in the case of acceptor bleaching or folding and unfolding. Only donor bleaching following acceptor bleaching would yield such a signal. Therefore, it is assumed, that most of the non-anti-correlated jumps are related to dark states of the dyes. In addition, the high number of Csp trajectories not showing any anti-correlated signal and the low number of anti-correlated jumps in all trajectories support the assumption that the trajectories do not contain much signal related to folding. This is in agreement with the transition maps, which show that bleaching is dominating the signal changes and only few events can be related to folding or unfolding.

It would be surprising if proteins had a completely different folding behavior if encapsulated in vesicles in comparison to ensemble experiments. A different folding behavior would also be in contradiction with previous work (44). However, the data presented here suggests that the proteins are destabilized in vesicles, as indicated by the observation that at GdmCl concentrations around the transition midpoint, almost only unfolded populations are present in both Csps in *E* histograms and transition maps of immobilization experiments. An explanation would be that proteins stick preferably to the lipid vesicles in the unfolded state, which could stabilize the unfolded state, but this hypothesis is in contradiction with fluorescence polarization measurement that show a rather narrow distribution, indication rotational freedom of the molecules. Another explanation would be that the fluorescence signal is rather dominated by dye photophysics. This is also controversial, because the same dyes were used as in a previous work (44). However, also in this previous work there is a possibility that intensity changes originating from dye photochemistry were included in the analysis.

### Additional transfer efficiency of $Pro_6$

It could not be clarified what the origin of the population with $E \approx 0.55$ for $Pro_6$ is. It is unlikely to be a result of conformational change, because proline peptides are present as a helix type II that has an extended conformation. It is known that proline peptides can undergo some bending, but in the case of $Pro_6$, $E$ of less than 0.95 means that the distance has increased, which is not possible because extension is limited by the bond length. Furthermore, populations with $E \approx 0.55$ are not present in single molecule free diffusion experiments, which suggests that this population is a result of altered fluorophore properties by photochemistry induced by long-term excitation or by immobilization. Additionally, for $Pro_{20}$, a worse peak separation between donor-only and proline peak than in free diffusion experiments indicates a higher contribution to lower $E$. Since longer distances are not possible, these contributions might be due to dark states.

### Possible explanations for dark states and outlook

It was recently found by Chung and coworkers (48) that Alexa Fluor 488 has a transient populated state with a red shifted emission. This was observed by measuring spectra of the emission of single dyes. It is highly probable that this red shifted state also appears in the measurements presented here. This would have two effects that both influence $E$. First, the donor has a lower intensity at the original donor (and measured) emission. Second, the overlap integral between donor emission and acceptor absorption spectra is changed, which results in a changed Förster radius. Third, more crosstalk into the acceptor channel is present. If the correction factor is not changed for the affected trajectory parts, additional populations with a larger $E$ could appear. This additional fluorescence intensity of the donor could contribute to additional complexity of the trajectories. However, in the work presented here, there are no additional populations in the transition maps or in the $E$ histograms observed. The peak separation in $E$ histograms is quite low, but in transition maps, additional populations should be visible.

The heterogeneity observed in the trajectories is most probably related to the unexplored photophysics or photochemistry of the dyes, such as the transient presence of dark states, the red shifted state of Alexa Fluor 488 or the formation of radicals (48, 55). A range of solution conditions and additives supposed to influence triplet and radical formation have been tested, but up to the end of the experiments, no positive effect could be identified. Since the project was terminated, some further additives that were able to increase lifetime and brightness of the dyes in diffusion and vesicle free immobilization experiments were tested in the group of Benjamin Schuler.

The problem could possibly also be addressed by choosing a different set of dyes, but it is also known that other dyes, as for example Cy3 and Cy5, have dark states. Additionally, both dyes are very hydrophobic, which could introduce some additional interaction with the vesicle surface. The choice of dyes had been limited by the excitation wavelength of the laser, but can be overcome now with the installation of fiber lasers, that allow extending the available wavelength range.

Currently, a number of experiments are ongoing to elucidate the source of photophysics interfering with our measurements. The role of dye-protein interactions that could induce dark states are addressed in experiments with DNA, labeled with the same dyes. Experiments with donor-only labeled protein demonstrated that the majority of trajectories showed continuous emission and final step-like bleaching. This could be checked for the acceptor by direct acceptor excitation, which is possible with a laser that has become available recently. Additionally, anti-correlated laser excitation (ALEX) of donor and acceptor during the entire observation would provide information of the state of the acceptor and allow the identification of trajectories or parts of them, in which both fluorophores are intact. Additional recording of fluorescence polarization could give information about the rotational freedom during the whole measurement.

### Summary

In summary, although fluorescence trajectories obtained from proteins encapsulated in vesicles were very heterogeneous and dominated by signal that was not obviously related to folding, but rather dependent on dye photophysics, it was possible to extract some folding signal related to folding. With a sophisticated automated analysis, trajectories or fragments of them were selected and further analyzed. Agreement was found in $E$ histograms of immobilization and diffusion experiments, but it was limited to the presence of one and two states for proline peptides and Csps, respectively and their dependence on GdmCl in the case of Csps. However, there was no agreement found in kinetic behavior to ensemble measurements. It was rather shown that the obtained kinetic signal cannot be related to folding, since it was not protein specific, but very similar folding and unfolding frequencies were measured for different proteins. Potential folding signal was recognizable only in very few cases. Further work will have to address the possible influence of dye photochemistry in more detail.

# Conclusion and outlook

This work consists of four parts: the first and the second part treat the collapse in protein folding and address the question, which interactions are dominant in the unfolded state; the third part treats the conformation of proteins bound to molecular chaperones, and in the fourth part, small proteins were immobilized to observe their individual folding behavior.

The focus of this work is the contribution of different interactions to protein collapse. The comparison of the temperature-induced collapse of a two-state folder and an intrinsically disordered protein (IDP) revealed that the collapse is not only driven by hydrophobic interactions, but that additional forces contribute to the compaction in the denatured state (42). There are indications from circular dichroism spectroscopy and simulations that there is secondary structure formation induced with increasing temperature that could lead to the collapse by forming non-native hydrogen bonds (63). It is still not clear which interactions are the dominant ones. Further steps could be to probe the temperature-collapsed state for local structure by nuclear magnetic resonance experiments. This might show whether specific interactions are participating, e.g., hydrogen bonds stabilize the collapsed state. Molecular simulations could also give further information about energetic contributions to unfolded state collapse.

The investigation of the denaturant-dependent collapse of IDPs with their high content of charged amino acids showed that the collapse is highly dependent on the average charge density and therefore dominated by electrostatic forces. A polyampholyte model (64) was applied to the data and was shown to describe the attractive and repulsive contributions that account for the final protein dimensions (65). These findings are important steps for the ultimate goal of quantifying the different contributions that cause proteins to fold. Additionally, information about the dimensions of IDPs and their strong salt dependence at low ionic strength is obtained. Consequences might be especially interesting for the function of IDPs.

In the third part of the work, the conformation and dynamics of a protein bound to a chaperone were investigated. By a combination of single molecule FRET and time-resolved fluorescence anisotropy decays, it was shown that in the case of rhodanese bound to the chaperone GroEL, fluorophore rotation is restricted, but that structural information still can be obtained from the FRET experiments by considering orientation factors (66). The results imply that the protein binds as a partially structured folding intermediate to the chaperone. This provides a general possibility to obtain distance information from systems with restrained dye rotation.

The aim of the fourth part of this thesis was to determine the folding and unfolding kinetics of single immobilized Csps and to compare them to the “average” folding rates of an ensemble of proteins. The results showed a high complexity in the trajectories, which is most probably due to dye photophysics and photochemistry. A comprehensive analysis method of trajectory data was developed and showed that the signal does not show solely protein folding and unfolding, but is dominated by the above mentioned dye effects. To detect folding and unfolding in these experiments, large improvements are needed. The next attempt of these kinds of experiments should include additives that protect the fluorophores from bleaching and temporal dark states (55-57). Further, the state of the acceptor dye should be checked, which can be done by alternating laser excitation (ALEX) (67) or pulsed interleaved excitation (PIE) (68). These methods are especially of advantage if large donor-acceptor distances are present which cannot be distinguished clearly from the donor-only population with conventional methods. Recently, a red-shifted state of the commonly used fluorophore Alexa Fluor 488 was discovered that also influences the energy transfer and leads to additional artificial populations in immobilization experiments (48). However, there were no indications found in this work for the described effect, but there are still many events that could not be explained and indicate further unknown photophysics. Future work will have to resolve such contaminating photochemical processes to utilize the full potential of single molecule fluorescence to investigate protein folding dynamics on time scales not accessible in free diffusion experiments.

# References

1. Pain RH (1996) *Mechanisms of Protein Folding* (Oxford University Press).
2. Bryngelson JD, Onuchic JN, Socci ND, & Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* 21(3):167-195.
3. Schuler B, Lipman EA, & Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419(6908):743-747.
4. Sherman E & Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc Natl Acad Sci USA* 103(31):11539-11543.
5. Laurence TA, Kong X, Jäger M, & Weiss S (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc Natl Acad Sci USA* 102(48):17348-17353.
6. Kuzmenkina EV, Heyes CD, & Nienhaus GU (2006) Single-molecule FRET study of denaturant induced unfolding of RNase H. *J Mol Biol* 357(1):313-324.
7. Deniz AA, Mukhopadhyay S, & Lemke EA (2008) Single-molecule biophysics: at the interface of biology, physics and chemistry. *J R Soc Interface* 5(18):15-45.
8. Schuler B & Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18(1):16-26.
9. Ziv G & Haran G (2009) Protein Folding, Protein Collapse, and Tanford's Transfer Model: Lessons from Single-Molecule FRET. *J Am Chem Soc* 131(8):2942-2947.
10. Perl D*, et al.* (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nat Struct Biol* 5(3):229-235.
11. Schuler B, Kremer W, Kalbitzer HR, & Jaenicke R (2002) Role of entropy in protein thermostability: folding kinetics of a hyperthermophilic cold shock protein at high temperatures using 19F NMR. *Biochemistry* 41(39):11670-11680.
12. Pauling L, Corey RB, & Branson HR (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci U S A* 37(4):205-211.
13. Schellman JA (1955) The stability of hydrogen-bonded peptide structures in aqueous solution. *C R Trav Lab Carlsberg Chim* 29(14-15):230-259.
14. Scholtz JM*, et al.* (1991) Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc Natl Acad Sci U S A* 88(7):2854-2858.
15. Lopez MM, Chin D-H, Baldwin RL, & Makhatadze GI (2002) The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. *Proc Natl Acad Sci U S A* 99(3):1298-1302.
16. Pace CN, Treviño S, Prabhakaran E, & Scholtz JM (2004) Protein structure, stability and solubility in water and other solvents. *Philos Trans R Soc Lond B Biol Sci* 359(1448):1225-1235.
17. Bolen DW & Rose GD (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* 77:339-362.
18. Honig B & Yang AS (1995) Free energy balance in protein folding. *Adv Protein Chem* 46:27-58.
19. Dill KA (1990) Dominant forces in protein folding. *Biochemistry* 29(31):7133-7155.
20. Honig B (1999) Protein folding: from the levinthal paradox to structure prediction. *J Mol Biol* 293(2):283-293.
21. Möglich A, Joder K, & Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103(33):12394-12399.
22. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1-63.

23. Klotz IM & Franzen JS (1962) Hydrogen bonds between model peptide groups in solution. *Journal of the American Chemical Society* 84(18):3461-3466.
24. Susi H & Ard JS (1969) Hydrophobic interactions and hydrogen bonding of epsilon-caprolactam in aqueous solution. *Journal of Physical Chemistry* 73(7):2440-2442.
25. Sadqi M, Lapidus LJ, & Munoz V (2003) How fast is protein hydrophobic collapse? *Proc Natl Acad Sci USA* 100(21):12117-12122.
26. Uversky VN, Gillespie JR, & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41(3):415-427.
27. Zhou H-X (2002) A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins. *Proc Natl Acad Sci USA* 99(6):3569-3574.
28. Hofmann H, Golbik RP, Ott M, Hübner CG, & Ulbrich-Hofmann R (2008) Coulomb forces control the density of the collapsed unfolded state of barstar. *J Mol Biol* 376(2):597-605.
29. Rose GD, Fleming PJ, Banavar JR, & Maritan A (2006) A backbone-based theory of protein folding. *Proc Natl Acad Sci USA* 103(45):16623-16633.
30. Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105-132.
31. Schulz GE (1979) *Nucleotide binding proteins. In: Molecular Mechanism of Biological Recognition* (Elsevier/North-Holland Biomedical Press).
32. Kriwacki RW, Hengst L, Tennant L, Reed SI, & Wright PE (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci USA* 93(21):11504-11509.
33. Dunker AK*, et al.* (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*:473-484.
34. Dunker AK*, et al.* (2001) Intrinsically disordered protein. *J Mol Graph Model* 19(1):26-59.
35. Wright PE & Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1):31-38.
36. Bustos DM & Iglesias AA (2006) Intrinsic disorder is a key characteristic in partners that bind 14-3-3 proteins. *Proteins* 63(1):35-42.
37. Dyson HJ & Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3):197-208.
38. Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293(2):321-331.
39. Glickman MH & Ciechanover A (2002) The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. (Translated from English) *Physiol Rev* 82(2):373-428 (in English).
40. Hoffmann A*, et al.* (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc Natl Acad Sci USA* 104(1):105-110.
41. Rüegger S (2008) Ensemble and Single-Molecule Fluorescence-Spectroscopic Analysis of Fragments of the Cold Shock Protein from the Hyperthermophilic Bacterium Thermotoga maritima. (Zürich University).
42. Nettels D*, et al.* (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc Natl Acad Sci U S A*.
43. Rhoades E, Gussakovsky E, & Haran G (2003) Watching proteins fold one molecule at a time. *Proc Natl Acad Sci USA* 100(6):3197-3202.
44. Rhoades E, Cohen M, Schuler B, & Haran G (2004) Two-state folding observed in individual protein molecules. *J Am Chem Soc* 126(45):14686-14687.
45. Boukobza E, Sonnenfeld A, & Haran G (2001) Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy. *JOURNAL OF PHYSICAL CHEMISTRY B* 105(48):12165-12170.
46. Okumus B, Wilson TJ, Lilley DMJ, & Ha T (2004) Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys J* 87(4):2798-2806.

47. Kuzmenkina EV, Heyes CD, & Nienhaus GU (2005) Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc Natl Acad Sci USA* 102(43):15471-15476.
48. Chung HS, Louis JM, & Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci U S A* 106(29):11837-11844.
49. Fernandez JM & Li H (2004) Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* 303(5664):1674-1678.
50. Camacho CJ & Thirumalai D (1993) Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA* 90(13):6369-6372.
51. Hohng S, Joo C, & Ha T (2004) Single-molecule three-color FRET. *Biophys J* 87(2):1328-1337.
52. Lee NK*, et al.* (2007) Three-color alternating-laser excitation of single molecules: monitoring multiple interactions and distances. *Biophys J* 92(1):303-312.
53. Clamme J-P & Deniz AA (2005) Three-color single-molecule fluorescence resonance energy transfer. *Chemphyschem* 6(1):74-77.
54. Heilemann M*, et al.* (2004) Multistep energy transfer in single molecular photonic wires. *J Am Chem Soc* 126(21):6514-6515.
55. Aitken CE, Marshall RA, & Puglisi J (2007) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J*.
56. Vogelsang J*, et al.* (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew Chem Int Ed Engl* 47(29):5465-5469.
57. Rasnik I, McKinney SA, & Ha T (2006) Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat Methods* 3(11):891-893.
58. Santoro MM & Bolen DW (1988) Unfolding Free-Energy Changes Determined by the Linear Extrapolation Method .1. Unfolding of Phenylmethanesulfonyl Alpha-Chymotrypsin Using Different Denaturants. (Translated from English) *Biochemistry* 27(21):8063-8068 (in English).
59. Fersht A (1999) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* ((New York: W.H.Freeman and Co.)).
60. Schuler B (2007) *From: Methods in Molecular Biology, vol. 350: Protein Folding Protocols* (Humana Press Inc., Totowa, NJ).
61. Ha T, Glass J, Enderle T, Chemla DS, & Weiss S (1998) Hindered rotational diffusion and rotational jumps of single molecules. (Translated from English) *Physical Review Letters* 80(10):2093-2096 (in English).
62. Schuler B, Lipman EA, Steinbach PJ, Kumke M, & Eaton WA (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *Proc Natl Acad Sci USA* 102(8):2754-2759.
63. Yang WY, Larios E, & Gruebele M (2003) On the extended beta-conformation propensity of polypeptides at high temperature. *J Am Chem Soc* 125(52):16220-16227.
64. Higgs PG & Joanny JF (1991) Theory Of Polyampholyte Solutions. *JOURNAL OF CHEMICAL PHYSICS* 94(2):1543-1554.
65. Müller-Späth S*, et al.* (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. (Translated from Eng) *Proc Natl Acad Sci U S A* (in Eng).
66. Hillger F*, et al.* (2008) Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew Chem Int Ed Engl* 47(33):6184-6188.
67. Kapanidis AN*, et al.* (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci USA* 101(24):8936-8941.
68. Müller BK, Zaychikov E, Brauchle C, & Lamb DC (2005) Pulsed interleaved excitation. (Translated from English) *Biophysical Journal* 89(5):3508-3522 (in English).

# Abbreviations

| | |
|---|---|
| AFM | atomic force microscopy |
| CD | circular dichroism |
| Csp*Bc* | cold shock protein from *Bacillus caldolyticus* |
| Csp*Tm* | cold shock protein from *Thermotoga maritima* |
| Csp | cold shock protein |
| Csps | cold shock proteins |
| DMSO | dimethyl sulfoxide |
| DTT | dithiothreithol |
| EDTA | ethylenediaminetetraacetic acid |
| EMCCD | electron multiplying charge coupled device |
| FRET | Förster resonance energy transfer |
| GdmCl | guanidinium chloride |
| GroEL/ES | bacterial molecular chaperone GroEL/ES |
| Hsp60 | heat shock protein 60 |
| IPTG | isopropyl-β-D-1-thiogalactopyranoside |
| LB | Luria-Bertani growth medium |
| LUVs | large unilamellar vesicles |
| OD | optical density |
| Pro6 | polyproline peptide with 6 prolines |
| Pro20 | polyproline peptide with 20 prolines |
| TIRF | total internal reflection microscopy |

# Acknowledgements

Many people supported me and enabled this work.

I like to thank Prof. Ben Schuler, for the possibility to work in his group and for offering me and initiating highly exciting projects for my thesis, for his constant support, and for providing excellent research facilities in his lab.

I thank Prof. Andreas Plückthun and PD Ilian Jelezarov for being in my PhD committee and Prof. Amedeo Caflisch for reviewing my thesis.

I like to thank the people who provided scientific support to make this work possible.

I thank Andrea Soranno for numerous discussions and the support with the data analysis. I thank Daniel Nettels for providing the software to analyse the data, for his support with Mathematica programming and the single molecule instrumentation. I thank Verena Hirschfeld for her help during protein purification optimization and sample preparation.

I thank all members of the Schuler group for their advice in experimental procedures, spontaneous help, and providing the great working atmosphere. I am very grateful to my colleagues Andrea Soranno, Daniel Nettels, Armin Hoffmann, Dominik Hänni, Verena Hirschfeld, Hagen Hofmann, Philipp Schütz, Frank Küster, Frank Hillger, Luc Reymond, René Wuttke, Bengt Wunderlich, Jennifer Clark, Ruth Kellner, Andrea Holla, and Stephan Benke. I like to thank Sara Züger for discussion and tips in molecular biology matters.

I like to thank Robert Seckler from Potsdam University for providing the possibility to start my thesis in his lab. I also like to thank the people in the Seckler lab for providing a great working atmosphere and facilitating my start in the lab.

I cordially thank my family and friends for their constant support and their interest in what I am doing. A special thank you goes to my parents Rosemarie und Wolfgang for giving me the possibility to study, for their encouragement, and their ongoing support. I like to thank Tomek for being with me.

# Lebenslauf

## Persönliche Daten

MÜLLER-SPÄTH (geborene GEISTER)
Sonja
geboren am 17. 09. 1976 in Braunschweig
Deutschland

## Ausbildung

| | |
|---|---|
| 8/89 – 6/96 | Lessinggymnasium Braunschweig-Wenden mit Erreichen der allgemeinen Hochschulreife |
| 9/96 - 2/97 | Studium der Elektrotechnik an der Fachhochschule Wolfenbüttel |
| 4/97 - 9/99 | Studium der Biologie an der Johannes Gutenberg-Universität in Mainz |
| 3/99 | Vordiplom mit Prüfungen in Botanik, Zoologie, Chemie und Physik |
| 9/99 - 6/00 | Einjähriges Auslandsstudium der Biochemie an der Cardiff University of Wales |
| 10/00 - 09/03 | Studium der Biologie mit Schwerpunkten in Botanik, Zoologie, Physiologischer Chemie und Informatik. |
| 30 September 03 | Erlangen des Diploms im Fach Biologie mit der Diplomarbeit „Stabilität des Lichtsammlerkomplexes LHCIIb, Mutationsanalyse der luminalen Schleife“ am Institut für Allgemeine Botanik der Johannes Gutenberg-Universität Mainz unter Betreuung von Prof. Dr. H. Paulsen |
| 10/03 - 8/04 | Promotion am Institut für Biologie und Biochemie der Universität Potsdam im Fach Physikalische Biochemie unter der Betreuung von Dr. B. Schuler |
| Seit 8/04 | Fortsetzung der Promotion als Doktorandin am Biochemischen Institut der Universität Zürich unter der Leitung von Prof. Dr. B. Schuler |